UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE ENGENHARIA DE BAURU
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

CAIO CÉSAR OBA RAMOS

DESENVOLVIMENTO DE FERRAMENTAS COMPUTACIONAIS INTELIGENTES
PARA IDENTIFICAÇÃO DE PERDAS COMERCIAIS EM SISTEMAS DE ENERGIA

BAURU

2010

CAIO CÉSAR OBA RAMOS

DESENVOLVIMENTO DE FERRAMENTAS COMPUTACIONAIS INTELIGENTES
PARA IDENTIFICAÇÃO DE PERDAS COMERCIAIS EM SISTEMAS DE ENERGIA

Dissertação apresentada como requisito á
obtenção do título de mestre em
Engenharia Elétrica, pelo Programa de
Pós-Graduação em Engenharia Elétrica,
da Faculdade de Engenharia de Bauru, da
Universidade Estadual Paulista.

Orientador: Prof. Dr. André Nunes de Souza

Co-orientador: Prof. Dr. João Paulo Papa

BAURU

2010

Ramos, Caio César Oba.

Desenvolvimento de Ferramentas Computacionais
Inteligentes para Identificação de Perdas Comerciais
em Sistemas de Energia / Caio César Oba Ramos, 2010.
80 f.

Orientador: André Nunes de Souza

Dissertação (Mestrado)-Universidade Estadual
Paulista. Faculdade de Engenharia, Bauru, 2010

1. Fraudes de Energia. 2. Perdas Comerciais. 3.
Técnicas Inteligentes. I. Universidade Estadual
Paulista. Faculdade de Engenharia. II. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE CAIO CESAR OBA RAMOS, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA, DO(A) FACULDADE DE ENGENHARIA DE BAURU.

Aos 02 dias do mês de março do ano de 2010, às 14:00 horas, no(a) ANFITEATRO DA PÓS-GRADUAÇÃO, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Prof. Dr. ANDRE NUNES DE SOUZA do(a) Departamento de Engenharia Elétrica / Faculdade de Engenharia de Bauru, Prof. Dr. PEDRO DA COSTA JÚNIOR do(a) Departamento de Engenharia Elétrica / Faculdade de Engenharia de Bauru, Prof. Dr. GERALDO FRANCISCO BURANI do(a) Departamento de Engenharia e Automação Elétrica / Escola Politécnica da USP, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de CAIO CESAR OBA RAMOS, intitulado "DESENVOLVIMENTO DE FERRAMENTAS COMPUTACIONAIS INTELIGENTES PARA IDENTIFICAÇÃO DE PERDAS COMERCIAIS EM SISTEMAS DE ENERGIA". Após a exposição, o discente foi arguido oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADO. Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.

Prof. Dr. ANDRE NUNES DE SOUZA

Prof. Dr. PEDRO DA COSTA JÚNIOR

Prof. Dr. GERALDO FRANCISCO BURANI

Dedico este trabalho com todo carinho
aos meus pais que me incentivaram e que me
deram total apoio para a concretização
desta conquista.

AGRADECIMENTOS

Agradeço primeiramente ao Prof. Dr. André Nunes de Souza que me orientou e foi muito paciente comigo durante todos esses anos, desde 2003, quando o procurei para fazer Iniciação Científica. Sua amizade foi um fator fundamental que contribuiu bastante para a minha produtividade científica.

Agradeço também o Prof. Dr. João Paulo Papa que me co-orientou, contribuindo muito no desenvolvimento deste trabalho e pela sua grande amizade desde os tempos de colégio.

Em especial quero agradecer todos os meus colegas de trabalho do LSISPOTI, que considero como amigos, Danilo Gastaldello, Lucas Pereira, Renato de Oliveira, Alexandre da Silva, João Alberto Zago Bevenuto, Kleber Ferraz e Ricardo Torrezan, pela contribuição, companhia e pela amizade em todos os momentos bons e difíceis que tivemos para cumprir todos os cronogramas.

Quero agradecer todos os professores da Pós-Graduação da Engenharia Elétrica da Faculdade de Engenharia de Bauru pelo conhecimento que adquiri.

Agradeço a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo incentivo e apoio financeiro dado para a execução deste trabalho.

E por fim agradeço a toda minha família, em especial aos meus pais, Eunice Oba e Alcides de Amorim Ramos, pela compreensão, incentivo, apoio e paciência que tiveram comigo durante o desenvolvimento deste trabalho.

RESUMO

A detecção de fraudes nos sistemas de energia provocados por consumidores ilegais é o principal alvo em estudos de perdas comerciais pelas empresas de energia. Comumente usadas entre as técnicas de reconhecimento de padrões, as Redes Neurais Artificiais e as Máquinas de Vetores de Suporte têm sido aplicadas para a identificação de fraudes de maneira automática, entretanto essas técnicas sofrem com a demora na convergência e no alto peso computacional. Este trabalho introduziu o classificador Floresta de Caminhos Ótimos para um rápido reconhecimento das perdas comerciais, que tem demonstrado ser superior às demais técnicas inteligentes, como as Redes Neurais e as Máquinas de Vetores de Suporte, sendo muito mais rápido. Neste trabalho, também foram apresentadas comparações entre esses classificadores.

Palavras-chave: Fraudes de Energia, Perdas Comerciais, Técnicas Inteligentes.

ABSTRACT

Fraud detection in energy systems by illegal consumers is the most actively pursued study in non-technical losses by electric power companies. Commonly used supervised pattern recognition techniques, such as Artificial Neural Networks and Support Vector Machines have been applied for automatic frauds identification, however they suffer from slow convergence and high computational burden. This work introduced the Optimum-Path Forest classifier for a fast non-technical losses recognition, which has been demonstrated to be superior than Neural Networks and Support Vector Machines, but much faster. Comparisons among these classifiers are also presented.

Keywords: Energy Frauds, Commercial Losses, Intelligent Techniques.

LISTA DE ILUSTRAÇÕES

Figura 3.1. Arquitetura bidimensional típica de uma rede de Kohonen.....	24
Figura 3.2. Modelo simplificado do Perceptron.....	27
Figura 3.3. Representação básica de uma rede <i>feedforward</i> (multicamadas).	28
Figura 3.4. Um problema de classificação binária cujo propósito é separar as bolas (o) dos diamantes (♦) onde o hiperplano ótimo é mostrado por uma linha sólida [19].....	33
Figura 3.5. Aplicação do truque do núcleo nas SVM, mapeando os dados de treinamento (“ <i>input space</i> ”) em um espaço de características (“ <i>feature space</i> ”).....	36
Figura 3.6. (a) Grafo completo. (b) MST do grafo completo. (c) Protótipos escolhidos. (d) Floresta de caminhos ótimos resultante. (e) Uma amostra de teste (triângulo). (f) O caminho ótimo do protótipo.	46
Figura 3.7. Processo de classificação do k-NN.	48
Figura 5.1. Gráfico de interpretação das informações de cada perfil de consumidor.	57
Figura 5.2. Esquema geral da arquitetura inteligente proposta.	59
Figura 6.1. Robustez dos classificadores a respeito de diferentes porcentagens de tamanho do conjunto de treinamento e teste (Bt).....	62
Figura 6.2. Robustez dos classificadores a respeito de diferentes porcentagens de conjunto de treinamento e teste (Bi).....	66
Figura 6.3. Robustez dos classificadores a respeito de diferentes porcentagens de conjunto de treinamento e teste (Bc).	66
Figura 6.4. Interface gráfica para o software desenvolvido.	67
Figura 6.5. Resultado da classificação dos consumidores feita pelo software desenvolvido...	69

LISTA DE TABELAS

Tabela 2.1. Perdas no Cenário Brasileiro	18
Tabela 6.1. Médias de Exatidão e Médias de Tempo de Treinamento e de Classificação pelo OPF, SVM-RBF, SVM-LINEAR e ANN-MLP para a Base Bt	61
Tabela 6.2. Médias de Exatidão e Médias de Tempo de Treinamento e de Classificação pelo OPF, SVM-RBF, SVM-LINEAR, ANN-MLP e SOM para a Base Bi.....	64
Tabela 6.3. Médias de Exatidão e Médias de Tempo de Treinamento e de Classificação pelo OPF, SVM-RBF, SVM-LINEAR, ANN-MLP e SOM para a Base Bc	65
Tabela 6.4. Regras para a Formatação da Planilha com os Dados de Cada Perfil de Consumidores no Padrão do Excel.....	68

LISTA DE ABREVIATURAS E SIGLAS

ANEEL – Agência Nacional de Energia Elétrica

ANN – *Artificial Neural Networks* (Redes Neurais Artificiais)

ANN-MLP – Redes Neurais Artificiais com Perceptrons Multicamadas

Bc – Base de Dados de Consumidores Comerciais

Bi – Base de Dados de Consumidores Industriais

Bt – Base de Dados para Teste de Consumidores Industriais

CODI – Comitê de Distribuição

DM – *Data Mining* (Mineração de Dados)

D_{max} – Demanda Máxima ou Demanda Medida

ELM – *Extreme Learning Machines*

FANN – *Fast Artificial Neural Network Library* (Biblioteca para ANN)

FC – Fator de Carga

FIFO – *First-In-First-Out*

GA – Algoritmos Genéticos

KDD – *Knowledge Discovery Databases*

KTU – Karush-Kuhn-Tucker

k-NN – k-Vizinhos Mais Próximos

kW – Quilowatts

LibSVM e LibLINEAR – Bibliotecas para Máquinas de Vetores de Suporte

LibOPF – Biblioteca para Floresta de Caminhos Ótimos

MST – *Minimum Spanning Tree* (Árvore Geradora Mínima)

NN – Vizinhos Mais Próximos

NLT – *Non-Technical Losses* (Perdas Não-Técnicas)

OPF – Floresta de Caminhos Ótimos

P_{inst} – Potência Instalada

SOM – Mapas de Auto-organização

SVM – *Support Vector Machines* (Máquinas de Vetores de Suporte)

SVM-LINEAR – Máquinas de Vetores de Suporte com Função Linear

SVM-RBF – Máquinas de Vetores de Suporte com Função Base Radial

TNB – Tenaga Nasional Berhad

SUMÁRIO

1	INTRODUÇÃO.....	11
1.1	Motivação	11
1.2	Objetivo	12
1.3	Estrutura	12
2	INVESTIGAÇÃO TÉCNICO-CIENTÍFICA	13
2.1	Perdas de Energia Elétrica.....	13
2.2	Definição de Fraude no Setor Elétrico	15
2.3	Procedimentos no Combate Contra as Perdas Comerciais.....	16
2.4	Perdas de Energia Elétrica no Cenário Brasileiro	17
3	ESTUDO E INVESTIGAÇÃO CIENTÍFICA SOBRE SISTEMAS INTELIGENTES ..	19
3.1	Introdução	19
3.2	Data Mining.....	19
3.3	Redes Neurais Artificiais.....	20
3.3.1	Mapas de Auto-Organização	22
3.3.2	Redes Neurais Perceptron Multicamadas	25
3.4	Máquinas de Vetores de Suporte	30
3.4.1	Introdução	30
3.4.2	Classificadores por Hiperplano	31
3.4.3	Nuclearização e Hiperplanos de Margem Suave.....	35
3.4.4	Considerações Adicionais	38
3.5	Floresta de Caminhos Ótimos.....	39
3.5.1	Introdução	39
3.5.2	Classificador por Floresta de Caminhos Ótimos	40
3.5.3	Fundamentação teórica	41
3.5.4	Treinamento.....	43
3.5.5	Classificação	45
3.6	K-Vizinhos Mais Próximos	46
3.7	Algoritmos Genéticos	49
4	ABORDAGENS INTELIGENTES PARA IDENTIFICAÇÃO DE CONSUMIDORES COM INSTALAÇÕES IRREGULARES	50
5	DIRETRIZ METODOLÓGICA DA PESQUISA.....	55
5.1	Introdução	55

5.2	Utilização da Base de Dados	55
5.2.1	Base de Dados do Primeiro Ensaio.....	58
5.2.2	Base de Dados do Segundo Ensaio.....	58
5.3	Arquitetura Inteligente.....	58
5.4	Bibliotecas Utilizadas na Implementação dos Classificadores.....	59
6	RESULTADOS E DISCUSSÕES.....	60
6.1	Introdução	60
6.2	Primeiro Ensaio (Base Bt)	60
6.2.1	Avaliação dos Classificadores	61
6.2.2	Robustez dos Classificadores	62
6.3	Segundo Ensaio (Bases Bi e Bc)	63
6.3.1	Avaliação dos Classificadores	64
6.3.2	Robustez dos Classificadores	65
6.4	Software de Aplicação Desenvolvido.....	67
6.5	Observações.....	70
7	CONCLUSÕES	71
7.1	Conclusões Gerais	71
7.2	Trabalhos Futuros	72
7.3	Publicações Durante a Pesquisa	72
	REFERÊNCIAS	74

1 INTRODUÇÃO

1.1 Motivação

Nos últimos anos, o sistema elétrico brasileiro vem passando por várias mudanças, consolidando-se o período de privatização das empresas geradoras, transmissoras e distribuidoras de energia elétrica, na qual introduziu um ambiente competitivo no cenário nacional. Os investimentos realizados pelas concessionárias têm como principal objetivo uma melhora significativa de seu desempenho financeiro e técnico, buscando produtividade, eficiência e rentabilidade. A fim de se obter uma melhor gestão com relação às perdas de energia, uma das formas de maximizar a energia disponível para comercialização é a concessionária de energia combater contra as perdas comerciais, principalmente as fraudes de energia [1], [2].

As perdas comerciais ou não-técnicas são aquelas associadas à comercialização da energia fornecida ao usuário e se referem à energia entregue e não faturada/cobrada, gerando uma perda no faturamento. Também são definidas como a diferença entre as perdas totais e as perdas técnicas, e estão principalmente relacionadas com as ligações clandestinas no sistema de distribuição [3].

Embora avanços nesta área tenham sido constatados nos últimos anos, principalmente no que diz respeito às diversas técnicas de medição de energia elétrica [4], torna-se cada vez mais necessária a pesquisa de métodos alternativos com grande flexibilidade e facilidade de adaptação para a contextualização do problema, como os modelos das técnicas computacionais com algoritmos inteligentes.

Entre as técnicas de inteligência computacional mais utilizadas para a solução de problemas estão as Redes Neurais Artificiais, Lógica Fuzzy, Algoritmos Genéticos, Máquinas de Vetores de Suporte, entre outras. As aplicações desses algoritmos inteligentes possibilitam o desenvolvimento de ferramentas computacionais utilizadas para a estimação e identificação de fraudes (perdas comerciais) em diversas empresas, analisando os dados de um determinado cliente e suas transações, sendo possível verificar se existe ocorrência de alguma operação suspeita de irregularidade.

1.2 Objetivo

Dentro deste contexto, situa-se o projeto de pesquisa intitulado “Desenvolvimento de Ferramentas Computacionais Inteligentes para Identificação de Perdas Comerciais em Sistemas de Energia”, que se caracteriza principalmente pela proposta de integrar elementos da Inteligência Computacional visando identificar problemas relacionados às fraudes em sistemas elétricos.

1.3 Estrutura

Esta dissertação está estruturada da seguinte maneira:

- O Capítulo 1 apresenta o objetivo e a motivação para o desenvolvimento deste trabalho;
- O Capítulo 2 define perdas comerciais e o que elas representam no cenário econômico e no setor elétrico;
- O Capítulo 3 descreve as técnicas de inteligência artificial mais utilizadas para a classificação de padrões, de maneira a minimizar os problemas do setor elétrico;
- O Capítulo 4 resume os principais estudos nacionais e internacionais relacionados com a utilização de algoritmos inteligentes para a identificação de perdas comerciais;
- O Capítulo 5 expõe a metodologia empregada neste trabalho para a identificação de perdas comerciais através de aplicações de técnicas inteligentes.
- O Capítulo 6 contém os resultados e discussões do trabalho desenvolvido; e
- O Capítulo 7 apresenta as conclusões, algumas sugestões de trabalhos futuros e os trabalhos que foram publicados durante o desenvolvimento deste trabalho.

2 INVESTIGAÇÃO TÉCNICO-CIENTÍFICA

2.1 Perdas de Energia Elétrica

As privatizações no setor de energia elétrica, a partir de 1995, determinando o fim do monopólio estatal, introduziram um ambiente competitivo no cenário nacional, forçando as empresas geradoras, transmissoras e distribuidoras de energia elétrica a procurar meios para aumentar suas receitas. Uma melhor gestão com relação às perdas de energia, que atualmente são da ordem de 16%, é uma das formas de maximizar a energia disponível para comercialização, alcançando, assim, produtividade, eficiência e rentabilidade [1].

A energia elétrica é um bem derivado da exploração dos recursos da natureza, portanto, o seu uso racional deve ser incentivado, sob pena de legarmos às gerações futuras, problemas sociais maiores do que a sua capacidade de solução. Quando se fala em racionalização do uso de um recurso energético, tem-se como consequência direta o adiamento de investimentos e, principalmente, a redução de impactos ambientais associados aos novos empreendimentos. Nesse sentido, é obrigação das empresas do setor elétrico em conjunto com a sociedade manter o controle das perdas de energia, racionalizando energia elétrica [2].

As empresas têm por direito uma gestão atuante, exemplo disso são seus investimentos que têm como principal objetivo uma melhora significativa do seu desempenho econômico–financeiro e técnico–operacional [2].

Para o maior entendimento de perdas elétricas, a seguir serão introduzidas algumas definições.

As Perdas de Energia Elétrica são constituídas pela diferença entre a energia gerada e/ou comprada por uma empresa de energia elétrica e a energia efetivamente faturada por ela. Tais perdas podem ser divididas em perdas técnicas e perdas comerciais:

a) Perdas Técnicas

As perdas técnicas são as energias perdidas no transporte, na transformação e nos equipamentos de medição, portanto inerente ao processo, isto é, um custo operacional normal para a concessionária. Essas perdas localizam-se tanto nos sistemas de geração, de transmissão, como também de distribuição de energia e ocorrem nos componentes agregados

ao sistema, tais como: condutores, equipamentos, conexões, transformadores, capacitores, reguladores de tensão, pára-raios, entre outros [3].

As perdas técnicas podem ser definidas e determinadas, por meio de simulações, medições e avaliações do sistema elétrico. Algumas metodologias abordam o tema, definindo mecanismos para a sua mensuração e controle. A redução de tais perdas depende, fundamentalmente, da tecnologia em utilização, da qualidade dos serviços de manutenção efetivados, da ampliação do sistema elétrico em concordância com a evolução do mercado consumidor e do modo de operação dos sistemas, onde alguns pontos chave de controle são: os níveis de carregamento dos condutores e transformadores, as demandas de energia reativa, os perfis da tensão, entre outros [1].

b) Perdas Comerciais

As perdas comerciais, o foco desta pesquisa, são aquelas associadas à comercialização da energia fornecida ao usuário final. Existe uma parcela de energia que é produzida, transportada e fornecida, no entanto, não é faturada, propiciando uma perda direta no faturamento [5], [6]. Elas são classificadas em administrativas e de intervenção ilícita.

As perdas comerciais de natureza administrativa possuem como causas:

- Sistema de medição incompleto, obsoleto ou sem calibração;
- Cadastro de consumidores ineficiente;
- Erro na leitura registrada pelo medidor ou no processo de informação dessa leitura; e
- Erro no processo de faturamento.

As intervenções ilícitas correspondem ao montante de energia não faturado, e estão relacionadas às intervenções que não são autorizadas pela concessionária, sendo realizadas por usuários ilícitos, com a finalidade de alterar o registro de energia elétrica ou o seu desvio [6].

As ligações clandestinas de energia elétrica são motivo de preocupação constante, tanto para as concessionárias de energia elétrica, como para as agências reguladoras. Neste contexto, os diversos tipos de fraudes podem afetar o poder público, que não recolhe os tributos e taxas de iluminação pública, e os consumidores em geral, que pagam pelo prejuízo rateado e deixam de receber benefícios sociais de obras que seriam realizadas pelo poder público. Outro enfoque também muito importante é que a minimização das perdas provocadas por estas fraudes representa a garantia de investimentos na qualidade do produto, na

manutenção do patrimônio da concessão e, principalmente, na possibilidade de expansão do atendimento com menor custo [5].

A necessidade de se combater as perdas comerciais se destaca pelo fato de que quanto maior elas são, menor é o faturamento e o lucro para a concessionária de energia e menor é o investimento de expansão para atender a demanda, pois as concessionárias de energia precisam cumprir todas as metas estabelecidas pela Agência Nacional de Energia Elétrica (ANEEL).

2.2 Definição de Fraude no Setor Elétrico

Conceitua-se como fraude no uso de energia elétrica, o ato de má-fé praticado contra a empresa fornecedora, o qual impede a correta medição e faturamento. Como exemplo de fraude, tem-se as ligações clandestinas, popularmente conhecidas como “gatos” [6].

A fraude ou furto de energia é estabelecido, quando ocorre, segundo um documento do CODI [7]:

- Desvio de energia elétrica em pontos localizados antes da medição;
- Adulteração dos mecanismos internos do medidor;
- Alteração dos registros de consumo;
- Interferência na chave de aferição;
- Interferência na fiação dos circuitos dos sistemas de medição indireta;
- Ligação ou religação à revelia: quando o consumidor por conta própria abre o medidor e faz sua instalação elétrica;
- Interferência ou dano provocado nos componentes elétricos; e
- Quaisquer outros atos provocados intencionalmente com o objetivo de diminuir o registro de energia elétrica efetivamente consumida.

Segundo Bettio [8], os tipos de fraudes mais comuns são:

a) Para consumidores de baixa tensão:

- Derivação antes da medição;
- Abertura da ponte de potencial do medidor de energia;

- Bloqueio do disco do medidor;
- Inversão dos fios de entrada com os de saída da unidade medidora;
- Descalibragem proposital do medidor;
- Desacoplamento do registrador; e
- Troca de engrenagens do registrador.

b) Para consumidores de alta tensão:

- Abertura da chave de aferição;
- Alargamento do pólo da chave de aferição, impedindo dessa forma o contato elétrico;
- Aplicação de verniz isolante na chave de aferição, impedindo a correta medição do consumo;
- Fios do circuito secundário dos transformadores de corrente (TC) de medição desencapados, formando pontos de contato entre condutores e reduzindo assim a corrente do medidor;
- Fios do circuito secundário dos transformadores de potencial (TP) interrompidos em determinados horários;
- Troca do TC por outro com relação de transformação superior, mantendo a placa original; e
- Curto-circuito provocado no primário dos TC's.

2.3 Procedimentos no Combate Contra as Perdas Comerciais

Com o objetivo de reduzir os valores das Perdas Comerciais, as empresas de energia elétrica atuam, geralmente em:

- Programas de inspeção: consiste na verificação da integridade do sistema de medição, detectando falhas nos equipamentos, fraudes e desvios de energia, erros de ligação e outros problemas que possam comprometer a medição de energia elétrica [9];
- Substituição de medidores: consiste na avaliação de lotes de medidores através de amostragem no campo, testes em laboratório e análise dos medidores retirados em campo. Além disso, a substituição de medidores com vida útil vencida ou com possíveis falhas técnicas;

- Regularização de focos possíveis de ligações clandestinas, tais como favelas, através de um programa de caráter regulatório, reduzindo conseqüentemente as perdas comerciais;
- Implementação de políticas comerciais: consiste em atendimentos para a comunidade sobre explicações, negociações e treinamentos sobre o consumo de energia elétrica; e
- Ações de eficiência energética com foco na diminuição das contas elétricas e uso efetivo da energia, servindo de incentivo aos consumidores para que não cometam as fraudes.

Uma das formas mais tradicionais de combate às perdas comerciais é a realização de inspeções nos consumidores. Entretanto, a seleção de quais consumidores que devem ser inspecionados é uma tarefa árdua para os especialistas no assunto. As distribuidoras geralmente empregam um conjunto de metodologias heurísticas para identificar os clientes de baixa tensão suspeitos de estarem cometendo algum tipo de irregularidade. Todavia, a média de acertos dessas metodologias ainda é bastante inferior ao desejado, acarretando prejuízos elevados para as distribuidoras brasileiras. Logo, a maioria dos processos adotados não é eficiente [10].

2.4 Perdas de Energia Elétrica no Cenário Brasileiro

De acordo com a ANEEL, as maiores perdas no cenário brasileiro atual ocorrem na região norte, onde a implementação de procedimentos contra as perdas técnicas e comerciais não são fáceis de executar, devido ao difícil acesso e a grande extensão territorial que a concessionária de energia atua, e as menores perdas ocorrem na região sul, um panorama oposto do que acontece na região norte, como pode ser visto na Tabela 2.1. Por exemplo, na região norte tem-se 18,4% de perdas totais, sendo 25% delas perdas comerciais, enquanto que na região sul, tem-se apenas 1,4% de perdas totais, sendo 10,3% delas perdas comerciais, evidenciando o contraste dos valores das perdas nestas regiões.

Através da literatura foi possível observar que as perdas em sistemas de energia é um problema que as empresas de energia enfrentam há bastante tempo.

Tabela 2.1. Perdas no Cenário Brasileiro

Região	Perdas Comerciais	Perdas Totais
Norte	25,0%	18,4%
Nordeste	19,9%	11,0%
Centro-Oeste	14,0%	1,8%
Sudeste	13,9%	3,0%
Sul	10,3%	1,4%

(Fonte: ANEEL)

A grande quantidade de ligações clandestinas existentes no Brasil e os números com relação à quantidade de perdas comerciais dão uma idéia da magnitude do problema e do grau de dificuldade para equacioná-lo [6].

Algumas empresas sofrem mais com os desvios de energia motivados pelo grande número de favelas (Rio de Janeiro, São Paulo, Bahia, Pernambuco) e outras são mais visadas pelas fraudes na própria unidade de medição (Paraná, Santa Catarina, Minas Gerais, Rio Grande do Sul). Além disso, existe um número considerável de irregularidades na medição que não constituem fraude, porém afetam diretamente a arrecadação da empresa [6].

Para reverter a atual situação, existem vários pontos que podem servir de início para se combater este grande nível de perdas comerciais, tais como o controle do faturamento da empresa, o controle de qualidade dos dados de cadastro dos consumidores verificando em campo qualquer mudança brusca em menor tempo possível e realizações de inspeções para detecção das irregularidades feitas por equipes especializadas e bem treinadas.

A base de dados é importante não apenas para examinar o faturamento, mas também é crucial para descobrir fraudes de energia, já que mantêm informações importantes tal como a classe dos consumidores (por exemplo: residencial, comercial, industrial), o endereço, aspectos técnicos da instalação elétrica (por exemplo: o número das fases), o transformador da distribuição a que o consumidor está conectado, histórico do consumidor (inadimplente ou não, possível caso de fraude), etc. [11].

3 ESTUDO E INVESTIGAÇÃO CIENTÍFICA SOBRE SISTEMAS INTELIGENTES

3.1 Introdução

Neste capítulo, é discutido como acontece o preparo dos bancos de dados através da mineração de dados (*data mining*) e os métodos aplicados para os classificadores inteligentes, através de uma aprendizagem supervisionada. Nesta pesquisa foram escolhidas para serem aplicadas: as Redes Neurais Artificiais (ANN), as Máquinas de Vetores de Suporte (SVM), que são as técnicas inteligentes mais utilizadas, com diversos trabalhos a nível internacional, e a Floresta de Caminhos Ótimos (OPF), que é uma técnica cuja aplicação no setor elétrico tem caráter inédito, para o desenvolvimento da metodologia proposta e a obtenção dos resultados devido à fácil adaptação na contextualização do problema e a grande flexibilidade dessas técnicas.

3.2 Data Mining

Numa época em que a maioria das operações e atividades é registrada computacionalmente e se acumulam em bases de dados, as empresas se deparam com o fato de possuírem enormes quantidades de informações, mas da qual não conseguem tirar proveito. Para analisar, explorar e interpretar tão grande quantidade de informação recorre-se ao auxílio da tecnologia, de forma a automatizar o processo de extração de informação [9].

O conceito de *Knowledge Discovery Databases* (KDD) pode ser definido como o processo de extração de informação a partir de dados registrados numa base de dados, extração essa não trivial de conhecimento implícito, previamente desconhecido, potencialmente útil e compreensível [12], [13].

Normalmente o conceito de *Data Mining* (DM) surge como um sinônimo de KDD. No entanto, o DM representa, apenas, uma das etapas do processo global KDD e consiste na utilização de modelos específicos para produzir padrões particulares, mostrando o seu valor diante os dados em análise. No seu todo, o processo poderá envolver as seguintes etapas:

- Limpeza de dados, de modo a remover ruídos e inconsistências;
- Integração de dados, onde dados provenientes de várias fontes e repositórios podem ser combinados e integrados em um *data warehouse*;
- Seleção de dados, onde é efetuada a escolha dos dados relevantes que se pretendem analisar;
- Transformação de dados, onde os dados são transformados e consolidados de acordo com o objetivo inicial e de modo a poderem ser aplicadas as técnicas de DM;
- *Data mining*, onde são aplicadas várias técnicas sobre os dados de modo a identificar padrões;
- Avaliação de padrões, para identificar os padrões que verdadeiramente representam conhecimento; e
- Apresentação do conhecimento adquirido, recorrendo a ferramentas de visualização com as quais são apresentados os resultados obtidos.

Em DM, uma tarefa corresponde a um tipo de problema de descoberta de conhecimento para o qual se pretende obter uma solução e fazendo uma abordagem de alto nível se consegue agrupá-las em dois grandes grupos: classificação e previsão.

Na classificação, o grande objetivo é encontrar fórmulas, regras ou padrões a partir de um conjunto de dados de treino, de modo a organizar esses dados em classes. Essas fórmulas ou regras irão depois ser aplicadas num conjunto de dados de teste. Na previsão, as operações são efetuadas sobre um conjunto de dados contínuos e ordenados, de modo a extrair um modelo preditivo que consiga prever valores futuros de outras variáveis de interesse [12], [13].

As técnicas de DM surgem na forma de algoritmos ou metodologias que são aplicadas na implementação das tarefas, como as árvores de decisão, redes neurais, entre outras.

3.3 Redes Neurais Artificiais

Redes Neurais Artificiais (ANN – *Artificial Neural Networks*) são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência. Uma ANN é

caracterizada pelo padrão de conexão (modelo conexionista) entre os neurônios (topologia), pelo método de determinação dos pesos e conexões (treinamento ou aprendizagem) e pela função de ativação responsável pela saída da rede. Esses processos comportam-se de maneira similar aos grupos de neurônios do cérebro humano que recebem e transmitem informações através dos dendritos e axônios, respectivamente [14].

Quando é apresentado à rede um conjunto de entradas e suas respectivas saídas, as quais estão representando o comportamento de um processo específico, a mesma, através do treinamento, é capaz de se auto-ajustar com o objetivo de mapear o relacionamento funcional entre as entradas e saídas. Por conseguinte, após a execução desse algoritmo de treinamento, a rede deve ser capaz de generalizar o comportamento do processo quando outras entradas, diferentes daquelas utilizadas no treinamento, são apresentadas. Tal característica é particularmente útil quando o relacionamento entre as entradas e saídas do processo analisado é não-linear, ou então, quando o relacionamento não é claramente definido de forma que a sua modelagem, por técnicas de identificação convencionais, torna-se muito difícil.

As redes neurais representam o desenvolvimento de sistemas computacionais capazes de reconhecer e classificar padrões, resolver problemas complexos, realizar processos indutivos e dedutivos, entre outros.

Em geral, a utilização de redes neurais, comparando-se com outras técnicas, possuem as seguintes vantagens [15]:

- Os relacionamentos funcionais entre os padrões de entrada e saída podem ser descobertos por uma rede neural, sendo que tais relacionamentos não precisam ser conhecidos ou descritos explicitamente;
- Nenhuma hipótese precisa ser feita sobre as distribuições estatísticas dos padrões de entrada; e
- As redes neurais são tolerantes a falhas no sentido que, mesmo quando alguns elementos ou conexões do processamento apresentam-se degradados, o desempenho da rede pode ser mantido robusto.

As arquiteturas de redes neurais artificiais utilizadas para modelar os sistemas neurais biológicos podem ser divididas em três categorias distintas. A primeira categoria de redes neurais, “*feedforward*” transforma conjuntos de sinais de entrada em conjuntos de sinais de saída, sendo que os parâmetros dessa transformação são ajustados de forma supervisionada, de acordo com os resultados desejados. Na segunda categoria de redes neurais, “*feedback*”, as informações de entrada definem o estado inicial de atividade do sistema de realimentação.

Depois de algumas transições de estado, atinge-se um estado que é considerado o resultado final da computação. Na terceira categoria, conhecida como mapas de auto-organização, as células (ou neurônios) vizinhas da rede neural interagem mutuamente, a fim de transformarem-se adaptativamente em detectores especializados de diferentes padrões. Nessa última categoria, o aprendizado da rede neural é realizado de forma não supervisionada [16].

3.3.1 Mapas de Auto-Organização

No modelo de redes neurais baseado em mapas de auto-organização (*Self-Organizing Maps* - SOM), proposto por Kohonen [16], a segregação espacial de diferentes respostas e suas organizações em subconjuntos relacionados topologicamente resultam em um alto grau de eficiência nas operações típicas de uma rede neural.

O modelo SOM implementa uma projeção não-linear de um espaço multidimensional X em um espaço bidimensional M , denominado mapa de auto-organização. Essa projeção, de modo análogo a maioria das projeções encontradas no cérebro, implementam mapeamentos topológicos, ou seja, elementos vizinhos em X são geralmente mapeados em elementos vizinhos em M .

O mapeamento topológico constitui uma importante característica do modelo de redes neurais proposta por Kohonen, pois permite encontrar características e outras abstrações do espaço multidimensional através de identificação de agrupamentos (“*clusters*”) no mapa bidimensional.

O mapeamento do espaço multidimensional no bidimensional é realizado por uma função $f: X \rightarrow M$, onde $X \subset \mathcal{H}^n$ e $M \subset \mathcal{H}^2$, que associa cada elemento $x \in X$ a um par $(i, j) \in M$. Os elementos m_{ij} do mapa M , assim como os dados de entrada, são vetores de tamanho n , que mantêm os pesos das ligações sinápticas da rede neural.

Apresentado um estímulo específico $x \in X$ a rede neural, o “nó vencedor” do mapa M para esse estímulo é determinado verificando-se qual dos elementos do mapa possui a menor distância ao estímulo, ou seja, o nó vencedor tem as coordenadas $(i, j) \in M$, tal que:

$$\|m_{ij} - x\| = \min\{\|m_{kl} - x\|, \quad \forall i, j, k, l = 0 \dots D\} \quad (3.1)$$

Onde D é o tamanho de M .

Portanto, um estímulo $x \in X$ apresentado à camada de entrada da rede neural é mapeado no elemento (i, j) da camada de saída da rede neural, ou seja, do mapa de auto-organização M , para o qual a distância entre esse estímulo e o vetor dos pesos das ligações sinápticas m_{ij} da rede neural seja mínima, comparando-se com os demais vetores de pesos das ligações sinápticas m_{kl} .

Nesse modelo de rede neural, os neurônios da camada de saída interagem lateralmente com seus vizinhos. Essas interações se dão na fase de aprendizado, quando os pesos do neurônio “vencedor” e dos seus vizinhos são ajustados, de tal modo que a neurônio vencedor tenha pesos idênticos aos valores do estímulo sendo apresentado à rede num determinado instante do treinamento, e os neurônios vizinhos tenham pesos semelhantes aos do neurônio vencedor, sendo que o grau de semelhança é ponderado de acordo com a distância entre o neurônio vizinho e o neurônio vencedor.

A interação do neurônio vencedor (i, j) com um neurônio vizinho (k, l) , num determinado instante t do treinamento, é definida por uma função $h_{kl}(t)$ da distância entre eles. Normalmente, a função $h_{kl}(t)$ é uma Gaussiana.

O aprendizado nesse modelo é dito não supervisionado porque o mapeamento do espaço multidimensional para o bidimensional é feito a despeito dos resultados esperados na saída do processamento, ao contrário dos outros modelos, onde os valores esperados e os valores obtidos pela rede para um dado estímulo de entrada influenciam no reajuste dos pesos das ligações sinápticas.

No modelo de redes neurais baseadas em mapas de auto-organização os pesos das ligações sinápticas são ajustadas da seguinte forma:

$$m_{kl}(t+1) = m_{kl}(t) + (x - m_{kl}(t))h_{kl}(t) \quad (3.2)$$

Onde $h_{kl}(t) = \alpha(t) \exp\left[-\frac{\|r_{kl} - r_{ij}\|^2}{\delta(t)^2}\right]$, (i, j) são as coordenadas do nó vencedor para o estímulo x , $\|r_{kl} - r_{ij}\|$ é a distância espacial entre o nó vencedor e o nó (k, l) e $\delta(t)$ é uma função do tempo de treinamento t , que determina a variância da Gaussiana h_{kl} .

O tamanho da vizinhança $N_{ij}(t)$ decresce de acordo com o número de iterações T da fase de treinamento. No início do treinamento, isto é, quando $t=0$, $N_{ij}(t)$ inclui toda a rede, e no final, quando $t=T$, $N_{ij}(t)$ contém somente o nó vencedor (i, j) . A seguinte função linear

pode ser adotada para determinar os ajustes nas extensões da vizinhança: $N_{ij}(t) = D [1 - t/T]$, para $t = 0, \dots, T$.

De forma análoga, as funções $\delta(t)$ e $\alpha(t)$, podem ser reajustadas em cada iteração através das funções lineares $\delta(t) = 2 - t/T$ e $\alpha(t) = 1 - t/T$, para $t = 1, \dots, T$, respectivamente. Após a fase de treinamento, os nós da rede são rotulados de acordo com a classe de estímulos para os quais eles apresentam respostas mais intensas. Se essa fase for supervisionada, então se deve apresentar para a rede conjuntos de estímulos para os quais se conhece a classificação. Para cada classe de estímulos conhecidos verificam-se quais são os nós vencedores no mapa de auto-organização. Esses nós são rotulados com o rótulo associado àquela classe de estímulos. Se essa fase não for supervisionada, é necessário fazer uma análise de agrupamentos no mapa de auto-organização.

Um modelo da arquitetura bidimensional típica da rede de Kohonen é apresentado na Figura 3.1.

Figura 3.1. Arquitetura bidimensional típica de uma rede de Kohonen.

3.3.2 Redes Neurais Perceptron Multicamadas

Perceptron

Durante as décadas de 50 e 60, uma grande revolução ocorreu no campo de pesquisa relacionado à teoria de reconhecimento de padrões com o aparecimento das chamadas máquinas que aprendem, os perceptrons [15].

Em sua metodologia mais básica, o perceptron aprende uma função de decisão linear que divide o conjunto de treinamento em dois subconjuntos, os quais são linearmente separáveis, em um número finito de passos iterativos. A resposta desse dispositivo básico baseia-se na soma ponderada de sua entrada, ou seja, pelas Equações (3.3) e (3.4), que é uma função de decisão linear em relação aos elementos do vetor de características.

$$d(x) = \phi(\xi) \quad (3.3)$$

$$\xi = \sum_{i=1}^n w_i x_i + w_{n+1} \quad (3.4)$$

Os coeficientes $w_j, j=1... n+1$ são os chamados pesos das conexões, sendo análogos às sinapses no sistema neural humano. Temos ainda que ϕ é a chamada função de ativação e ξ é o somatório da rede [14]. Dentre as várias escolhas para ϕ temos:

- Função limiar (threshod):

$$\phi(\xi) = \begin{cases} 1 & \text{se } \xi \geq 1, \\ 0 & \text{caso contrário} \end{cases} \quad (3.5)$$

- Função sigmoidal:

$$\phi(\xi) = \frac{1}{1 + \exp(-\xi)} \quad (3.6)$$

- Função identidade:

$$\phi(\xi) = \xi \quad (3.7)$$

O algoritmo de treinamento do perceptron utiliza uma função de ativação por limiarização muito semelhante à Equação (3.5), descrita por:

$$\phi(\xi) = \begin{cases} 1 & \text{se } \xi \geq 1, \\ -1 & \text{caso contrário} \end{cases} \quad (3.8)$$

Geometricamente, a Equação (3.9) ou (3.10) define um hiperplano no espaço n -dimensional de padrões.

$$\xi = \sum_{i=1}^n w_i x_i + w_{n+1} = 0 \quad (3.9)$$

$$\xi = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + w_{n+1} = 0 \quad (3.10)$$

Os n primeiros coeficientes estabelecem a orientação do hiperplano, enquanto o último coeficiente, w_{n+1} , também chamado de *bias*, é proporcional a distância perpendicular a origem do hiperplano.

Caso as duas classes apresentadas ao algoritmo de perceptron sejam linearmente separáveis em $\mathfrak{R}^n$, o mesmo converge em um número finito de passos. Entretanto, caso ambas as classes não obedeçam a tal critério, o algoritmo executara indefinidamente, não convergindo para a solução ótima.

O modelo simplificado do perceptron é apresentado na Figura 3.2.

Figura 3.2. Modelo simplificado do Perceptron.

Perceptron Multicamadas

Para o tratamento de problemas que possuem várias classes, independentemente de as mesmas serem separáveis ou não, o algoritmo de perceptron visto anteriormente é ineficiente, visto que o mesmo funciona adequadamente quando a tarefa de classificação consiste em apenas duas classes linearmente separáveis. Conectando vários perceptrons, podemos projetar uma estrutura chamada Perceptron Multicamadas, também conhecida como ANN-MLP (*Artificial Neural Networks – Multilayer Perceptron*) [15], a qual consiste de várias camadas de elementos computacionais idênticos (neurônios) dispostos de tal maneira que a saída de cada um deles alimente a entrada de cada neurônio da camada seguinte. O número de neurônios da primeira camada, denominada A , corresponde a N_A , sendo, frequentemente denotado $N_A = n$, ou seja, o número de elementos da primeira camada é igual à dimensionalidade do vetor de características (ou atributos). O número de neurônios da camada de saída, chamada Q , corresponde a $N_Q = M$, sendo M igual ao número de classes do problema de classificação. A rede neural reconhece um vetor de padrões x como pertencente à classe ω_m caso a m -ésima saída da rede possuir o maior valor dentre as $m - 1$ saídas restantes.

No caso da arquitetura citada anteriormente, a entrada de cada elemento em qualquer camada corresponde à soma ponderada das saídas da camada anterior. Denotando K a camada anterior a J , tem-se que a entrada de cada neurônio na camada J , denotada por I_j é dada pela Equação (3.11).

$$I_j = \sum_{k=1}^{N_k} w_{jk} O_k \quad (3.11)$$

Sendo que:

$$O_k = \phi(I_k) \quad (3.12)$$

Onde $j = 1, 2, \dots, N_J$, sendo N_J e N_K o número de elementos da camada J e K , respectivamente, e w_{jk} são os pesos que modificam as saídas O_k dos elementos da camada K .

Algumas convenções geralmente adotadas para tais estruturas de redes neurais são:

- A função de ativação utilizada na camada de entrada é a função identidade, dada pela Equação (3.7);
- Camadas não adjacentes não são conectadas diretamente; e
- Todos os neurônios das camadas intermediárias possuem a mesma função de ativação ϕ .

Sendo assim, este modelo de rede neural é chamado de *feedforward*, devido ao fato de tanto a camada de entrada quanto as intermediárias (ou escondidas) serem submetidas somente à camada mais alta, ou seja, a camada K e subalterna a camada $K-1$, e assim por diante.

Uma representação de um modelo básico do modelo *feedforward* pode se visto na Figura 3.3.

Figura 3.3. Representação básica de uma rede *feedforward* (multicamadas).

Treinamento por Retropropagação

O principal objetivo deste algoritmo é desenvolver uma regra de treinamento com o intuito de minimizar o erro quadrático total entre as saídas desejadas r_q e as saídas reais O_q dos nós em uma camada de saída Q , ou seja, minimizar a Equação (3.13).

$$E_Q = \frac{1}{2} \sum_{q=1}^{N_Q} (r_q - O_q)^2, \quad (3.13)$$

Onde N_Q é o número de nós da camada de saída Q .

Durante o treinamento com o algoritmo de retropropagação, primeiramente, um padrão é apresentado à camada de entrada da rede. A saída obtida pela última camada é comparada com a saída desejada. Caso a mesma não estiver correta, o erro é calculado através da Equação (3.13) e propagado a partir da camada de saída até a camada de entrada, sendo que os pesos das conexões são modificados através da regra delta generalizada conforme o erro é retropropagado [17].

Desta forma, a regra delta generalizada Δw_{qp} permite o ajuste dos pesos em cada uma das camadas de maneira a tentar minimizar a função de erro (3.13), ou seja, calcular o gradiente negativo de E_Q , dado pela Equação (3.14).

$$\Delta w_{qp} = -\alpha \frac{\partial E_Q}{\partial w_{qp}}, \quad (3.14)$$

Em que P precede a camada Q e α é um incremento positivo de correção.

Pode-se resumir e generalizar o procedimento de treinamento da seguinte maneira: para quaisquer duas camadas K e J , em que K precede imediatamente J , a Equação (3.14), através de algumas manipulações algébricas, pode ser re-escrita como a Equação (3.15).

$$\Delta w_{jk} = \alpha \delta_j O_k \quad (3.15)$$

Caso J seja uma camada de saída, δ_j é dado por:

$$\delta_j = (r_j - O_j) O'_j \quad (3.16)$$

Se J for uma camada interna e P a próxima camada, então δ_j é dado por:

$$\delta_j = O'_j \sum_{p=1}^{N_p} \delta_p w_{jp}, \quad (3.17)$$

Onde $j=1, 2, \dots, N_j$.

Desta forma, todo o processo de treinamento começa com um conjunto arbitrário de pesos da rede. Em seguida, a aplicação da regra delta generalizada em qualquer passo iterativo envolve duas etapas. Na primeira fase, um vetor de treinamento é apresentado à rede e propagado através das camadas da rede para o cálculo de O_j para cada nó. As saídas O_q dos nós da camada de saída são então comparadas com as respostas desejadas r_q para que os termos de erro δ_q sejam gerados. A segunda fase envolve uma passagem em direção à entrada da rede, retropropagação, durante a qual o sinal de erro apropriado é passado por cada nó e as mudanças correspondentes nos pesos são realizadas. Uma vez que o sistema tenha sido treinado, o mesmo passa a classificar os padrões utilizando os parâmetros estabelecidos durante a fase de treinamento descrita nesta seção [18].

3.4 Máquinas de Vetores de Suporte

3.4.1 Introdução

As Máquinas de Vetores de Suporte (SVM) - do inglês “*Support Vector Machines*” - são normalmente consideradas a primeira aplicação prática da teoria do aprendizado estatístico [19], [20]. Trata-se de uma área de pesquisa que oferece muitas opções para se trabalhar, sendo grande parte delas mais conceituais do que meramente práticas. Nos últimos anos seu escopo tem crescido significativamente tanto em termos de novos algoritmos quanto de um entendimento teórico mais aprofundado. Parte destes novos algoritmos se deve aos chamados métodos de nuclearização, uma proposta para solução de problemas de *aprendizado de máquina* cuja arquitetura tem se demonstrado capaz de lidar com questões relativas às bases desta teoria. Além disso, aplicações bem sucedidas de SVM demonstram que esta técnica não só possui uma fundamentação mais sólida do que as redes neurais

artificiais como também são capazes de substituí-las com desempenho melhor ou semelhante [19].

As SVM propõem resolver o problema de classificação de padrões assumindo ser possível separar as classes em um espaço de mais alta dimensão. Suponha uma situação na qual os dados não são linearmente separáveis. Tais amostras podem ser separadas em grupos usando curvas ou círculos como superfícies de decisão, porém encontrar tais limiares é uma tarefa custosa. A principal idéia de uma SVM é pré-processar os dados de tal forma que o problema de encontrar uma função discriminante não-linear seja transformado em um problema de encontrar um hiperplano, ou seja, mapear os dados que estão em uma dimensão qualquer para outra maior, tornando os mesmos linearmente separáveis. Isto é feito definindo um mapeamento que transforma o vetor de entrada em outro (usualmente maior) vetor. Espera-se que, escolhendo um mapeamento adequado, o novo conjunto de treinamento seja linearmente separável.

3.4.2 Classificadores por Hiperplano

Nos primórdios de seu estudo, Vapnik [20] considerou uma classe de hiperplanos em um espaço H com produto interno:

$$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0 \quad (3.18)$$

Onde $\mathbf{w} \in H$, $b \in \mathfrak{R}$, correspondendo com funções de decisão do tipo:

$$f(x) = \text{sgn}(\langle \mathbf{w}, \mathbf{x} \rangle + b) \quad (3.19)$$

Baseado em dois argumentos, ele então propôs o algoritmo de aprendizado denominado Retrato Generalizado (do inglês “*Generalized Portrait*”) para problemas separáveis por hiperplanos [20]:

1. Dentre todos os hiperplanos que separam os dados, existe apenas um *hiperplano ótimo* distinguido pela margem de máxima separação entre qualquer ponto de treinamento e este hiperplano. Esta é a solução de:

$$\begin{aligned} & \text{maximize} \\ & w \in H, b \in \Re \quad \min\{\|\mathbf{x} - \mathbf{x}_i\| \mid \mathbf{x} \in H, \langle \mathbf{w}, \mathbf{x} \rangle + b = 0, i=1, \dots, m\} \end{aligned} \quad (3.20)$$

2. A capacidade da classe de hiperplanos de separação decresce com o crescimento da margem.

Para construir tal hiperplano ótimo, é necessário resolver:

$$\begin{aligned} & \text{minimize} \\ & w \in H, b \in \Re \quad \tau(\mathbf{w}) = 0,5\|\mathbf{w}\|^2 \end{aligned} \quad (3.21)$$

Sujeito à

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 \quad \text{para todo} \quad i = 1, \dots, m \quad (3.22)$$

Com a restrição (3.22) garantindo que $f(x_i)$ será +1 para $y_i = +1$ e -1 para $y_i = -1$ e também fixando a escala de $\mathbf{w}$.

A razão por detrás da minimização de $\mathbf{w}$ em (3.21) pode ser interpretada da seguinte maneira: se $\|\mathbf{w}\|$ fosse 1, então o termo da esquerda de (3.22) seria igual à distância de x_i ao hiperplano. Em geral, é preciso dividir $y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b)$ por $\|\mathbf{w}\|$ para transformá-lo nesta distância. Sendo assim, se (3.22) for satisfeita para todos $i = 1, \dots, m$ com um $\mathbf{w}$ de tamanho mínimo, a margem será maximizada como um todo (ver Figura 3.4).

Figura 3.4. Um problema de classificação binária cujo propósito é separar as bolas (o) dos diamantes (♦) onde o hiperplano ótimo é mostrado por uma linha sólida [19].

A função τ em (3.21) é chamada de *função objetivo*, enquanto (3.22) representa as restrições de *desigualdade*. Juntas elas dão origem ao que se conhece como *problema de otimização restrita*, que por sua vez é resolvido através da introdução do *Lagrangiano*:

$$L(\mathbf{w}, b, \alpha) = 0,5\|\mathbf{w}\|^2 - \sum_{i=1}^m \alpha_i (y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1) \quad (3.23)$$

Onde $\alpha_i \geq 0$ são os multiplicadores de Lagrange e L tem que ser minimizada com relação às variáveis primais $\mathbf{w}$ e b e maximizada com relação às variáveis duais α_i . Em outras palavras, um ponto de sela deve ser encontrado. Como se pode notar, as restrições do problema original foram incorporadas no segundo termo do Lagrangiano.

Para o problema de otimização restrita, é possível observar que se a restrição (3.22) no segundo termo do Lagrangiano for violada com $\sum_{i=1}^m \alpha_i (y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) - 1) < 0$, L pode ser incrementado através do decrescimento do α_i correspondente. Ao mesmo tempo, $\mathbf{w}$ e b terão que mudar (reduzindo a margem) de tal forma que L decresça. Dado que o problema é separável, num dado momento as restrições serão finalmente satisfeitas. De forma análoga, podemos entender que para todas as restrições não atendidas precisamente como igualdades, seu α_i correspondente deve ser 0: este é o valor de α_i que maximiza L . Desta observação extrai-se as chamadas condições de complementaridade de Karush-Kuhn-Tucker (KKT) presente na teoria da otimização [20].

Considerando que num ponto de sela as derivadas de L com relação às variáveis primais devem desaparecer [19], [20]:

$$\frac{\partial}{\partial b} L(\mathbf{w}, b, \alpha) = 0 \quad \text{e} \quad \frac{\partial}{\partial \mathbf{w}} L(\mathbf{w}, b, \alpha) = 0 \quad (3.24)$$

Levando à:

$$\sum_{i=1}^m \alpha_i y_i = 0, \text{ e} \quad (3.25)$$

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \quad (3.26)$$

Conseqüentemente, o vetor-solução nada mais é que uma expansão (3.26) de um subconjunto dos padrões de treinamento, especificamente aqueles com α_i diferente de zero, que são os chamados vetores de suporte (SVs – do inglês “*Support Vectors*”).

Ao substituir (3.25) e (3.26) no Lagrangiano (3.23), elimina-se as variáveis primais $\mathbf{w}$ e b resultando no chamado *problema de otimização dual*, que é o tipo de problema que as SVM normalmente resolvem na prática:

$$\underset{\alpha \in \Re^m}{\text{maximize}} \quad W(\alpha) = \sum_{i=1}^m \alpha_i - 0,5 \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (3.27)$$

$$\text{Sujeito à } \alpha_i \geq 0 \quad \text{para todo } i = 1, \dots, m \quad \text{e} \quad \sum_{i=1}^m \alpha_i y_i = 0. \quad (3.28)$$

Ainda baseado em (3.26), a função-hiperplano de decisão (3.19) pode ser escrita como:

$$f(x) = \text{sgn}\left(\sum_{i=1}^m y_i \alpha_i \langle \mathbf{x}, \mathbf{x}_i \rangle + b\right) \quad (3.29)$$

Para b poder ser calculado, seguindo as condições KKT, todos os pontos de acordo com (3.30) são os vetores de suporte (SVs) situados na margem e, sendo assim, $\alpha_i > 0$.

$$\alpha_i [y_i (\langle \mathbf{x}_i, \mathbf{w} \rangle + b) - 1] = 0 \quad \text{para todo} \quad i = 1, \dots, m \quad (3.30)$$

Em tais casos, é possível demonstrar que:

$$\langle \mathbf{x}_i, \mathbf{w} \rangle + b = y_i \quad (3.31)$$

Dessa forma, o limiar pode ser obtido, por exemplo, por:

$$b = y_i - \langle \mathbf{x}_i, \mathbf{w} \rangle \quad (3.32)$$

Alternativamente, b também pode ser calculado a partir dos valores das variáveis duais α_i e α_j correspondentes.

3.4.3 Nuclearização e Hiperplanos de Margem Suave

Esta seção está relacionada com o mapeamento das entradas para um espaço de maior dimensionalidade e ao acréscimo de artifícios de tolerância para que as SVM possam lidar com problemas não-separáveis.

Para expressar as fórmulas a partir de padrões de entrada $\mathcal{X}$, emprega-se o produto interno dos vetores $\mathbf{x}, \mathbf{x}'$ em termos do núcleo k estimado pelos elementos de entrada x, x'

$$k(x, x') = \langle \mathbf{x}, \mathbf{x}' \rangle \quad (3.33)$$

Esta substituição, também conhecida como *truque do núcleo* (do inglês “*kernel trick*”), é usada para estender o conceito de classificação por hiperplanos para máquinas de vetores de suporte não-lineares. O truque do núcleo pode ser aplicado desde que todos os vetores de característica ocorram apenas em produtos internos, o que pode ser observado nas Equações (3.27) e (3.29). O vetor ponderado (3.26) torna-se então uma expansão no espaço de

características e, sendo assim, não mais corresponde ao seu respectivo vetor no espaço de entradas. A partir de (3.29), a função de decisão se torna:

$$f(x) = \text{sgn}\left(\sum_{i=1}^m y_i \alpha_i k(x, x_i) + b\right) \quad (3.34)$$

O problema de otimização (3.27) assume a forma:

$$\begin{aligned} &\text{maximize}_{\alpha \in \mathfrak{R}^m} W(\alpha) = \sum_{i=1}^m \alpha_i - 0,5 \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j k(x_i, x_j) \end{aligned} \quad (3.35)$$

Sujeito à $\alpha_i \geq 0$ para todo $i = 1, \dots, m$ e $\sum_{i=1}^m \alpha_i y_i = 0$.

A Figura 3.5 captura a idéia de mapear os dados do espaço de entradas para o espaço de características de maior dimensão através de Φ (função não-linear) e construir lá um hiperplano de máxima margem de separação. Apesar da fronteira de decisão ser linear no espaço de características, no espaço de entradas original ela ganha formas não-lineares.

Figura 3.5. Aplicação do truque do núcleo nas SVM, mapeando os dados de treinamento (“*input space*”) em um espaço de características (“*feature space*”).

Mesmo com as vantagens da nuclearização do problema, na prática, o hiperplano de separação ainda assim pode não existir. Isto pode acontecer, por exemplo, em um conjunto de dados com muito ruído onde haja sobreposição das classes. Para permitir que os exemplos que violam a Equação (3.22) possam ser considerados, as variáveis de afrouxamento ξ_i são introduzidas [19], [21], onde:

$$\xi_i \geq 0 \quad \text{para todo} \quad i = 1, \dots, m \quad (3.36)$$

Fazendo com que as restrições assumam a forma:

$$y_i (\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \quad \text{para todo} \quad i = 1, \dots, m \quad (3.37)$$

Desta maneira, um classificador de boa generalização pode ser obtido pelo controle tanto da capacidade (através de $\|\mathbf{w}\|$) quando da soma das variáveis de afrouxamento $\sum_i \xi_i$. É possível demonstrar que este somatório provê um limite superior no número de erros de treinamento [19].

Neste contexto, uma possível elaboração de tal classificador de *margem suave* é obtida pela minimização da função objetivo:

$$\pi(\mathbf{w}, \xi) = 0,5\|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \quad (3.38)$$

Sujeita às restrições (3.36) e (3.37), onde a constante $C > 0$ determina o equilíbrio entre a maximização da margem e a minimização dos erros de treinamento. Quanto maior for C , menor será a margem, menor o número de erros de treinamento e menor também a capacidade de generalização da máquina de aprendizado.

Incorporando o núcleo e reescrevendo o problema em termos dos multiplicadores de Lagrange novamente nos leva a maximização de (3.35), sujeita às restrições:

$$0 \leq \alpha_i \leq C \quad \text{para todo} \quad i = 1, \dots, m \quad e \quad \sum_{i=1}^m \alpha_i y_i = 0 \quad (3.39)$$

A única diferença deste caso para o caso separável está no limite superior C dos multiplicadores de Lagrange α_i . Sendo assim, a influência dos padrões individuais (que podem ser valores discrepantes) se torna limitada. A solução em (3.34) não muda e o limiar b pode ser calculado pela exploração do fato de que para todos os SVs x_i com $\alpha_i < C$, a variável de afrouxamento ξ é zero e, assim sendo:

$$\sum_{j=1}^m \alpha_j y_j k(x_i, x_j) + b = y_i \quad (3.40)$$

Analisando geometricamente, escolher b significa deslocar o hiperplano, ao passo que (3.40) determina ser necessário situá-lo de tal forma que todos os SVs com variável de afrouxamento igual a zero estejam nas linhas ± 1 (conforme Figura 3.4).

A função de núcleo apresentada como produto interno em (3.33) é normalmente chamada de função de *Núcleo Linear* e, dadas algumas restrições [19], pode ser substituída por outras funções. Outra função de mapeamento bastante conhecida é a gaussiana chamada *Função de Base Radial* (do inglês “*Radial Basis Function*” - RBF):

$$k(x, x') = \exp\left(\frac{-\|x - x'\|^2}{2\sigma^2}\right) \quad (3.41)$$

Onde $\sigma > 0$ é um parâmetro da função de núcleo. Em função da variável de ajuste C , este tipo de SVM é comumente chamado de C-SVC e representa a classificação por SVM em sua forma original [22].

3.4.4 Considerações Adicionais

Como foi possível observar, as SVM são fundamentalmente máquinas de classificação binária. No entanto, a maioria dos problemas reais a serem otimizados requerem a classificação de mais de duas classes. Uma questão que emerge desta constatação é sobre como lidar com a propriedade multiclasse inerente a muitos problemas em um classificador binário por definição. Três estratégias são comumente adotadas para emprego de SVM em problemas com k classes, ou seja, multiclass [23]-[26]:

- **Uma contra o resto:** na abordagem uma-contra-o-resto, k SVM são treinadas, onde cada uma delas separa uma determinada classe de todas as demais.

- **Um contra um:** técnica também conhecida como classificação aos pares. Nesta abordagem de classificação, $k(k - 1)/2$ máquinas são treinadas, sendo que cada uma destas SVM separa um par de classes. Assim, temos todas as possíveis combinações de classes duas a duas representadas, individualmente, por uma SVM. Os classificadores são então arranjados como nós de uma árvore, onde cada nó representa uma SVM. Tanto a busca de cima para baixo quanto de baixo para cima podem ser adotadas, sendo esta última análoga ao processo de eliminação de equipes em um campeonato esportivo.
- **Espaço de diferenças:** para transformar problemas multiclasse em binários, um novo espaço de representações chamado *espaço de diferenças* foi proposto em [21]. Através da modelagem de dissimilaridades, um problema com k classes é transformado em um problema com os conjuntos *diferenças intraclasse* e *diferenças interclasses*. Quanto mais parecida do conjunto intraclasse for a diferença entre um exemplo conhecido e um desconhecido, mais provável os dois serem da mesma classe. Por outro lado, quanto mais parecida do conjunto interclasse for a diferença entre um exemplo conhecido e um desconhecido, mais provável deles serem de classes distintas.

Uma biblioteca de código aberto bastante conhecida que implementa SVM nas mais variadas formas é a chamada LibSVM [27]. Trata-se de um repositório de código aberto cujo objetivo é ajudar as pessoas a usarem facilmente SVM. Além de prover quatro diferentes tipos de função de núcleo, ela oferece SVM não só para classificação, mas também para “clusterização” e regressão. Através de uma interface por linhas de comando, é possível manipular os conjuntos de treinamento e teste, treinar SVM de diferentes maneiras, buscar por parâmetros ótimos das funções de núcleo e executar as classificações.

3.5 Floresta de Caminhos Ótimos

3.5.1 Introdução

É apresentada nesta seção uma técnica de classificação supervisionada de padrões baseadas em grafos, chamada Floresta de Caminhos Ótimos (*Optimum-Path Forest* - OPF).

Abordagens baseadas em OPF tratam as amostras como sendo os nós de um grafo, sendo os arcos definidos por uma relação de adjacência e ponderados por alguma métrica de distância aplicada em seus vetores de atributos, e diferem dos métodos tradicionais por não utilizar a idéia de geometria do espaço de atributos conseguindo, assim, melhores resultados em bases com *outliers* (amostras de uma determinada classe que estão presentes em uma região de outra classe no espaço de atributos) e sobreposição de classes. Duas abordagens supervisionadas já foram propostas atualmente, as quais diferem tanto na relação de adjacência e na função de valor de caminho utilizadas, quanto na maneira de encontrar os protótipos. A primeira delas utiliza como relação de adjacência o grafo completo e busca como protótipos as amostras que pertencem à intersecção entre as classes no conjunto de treinamento [28]. A outra metodologia desenvolvida utiliza um grafo k-NN (*k-Nearest Neighbor*, ou seja, k-vizinhos mais próximos) e encontra os protótipos como sendo os máximos regionais ou amostras de cada classe na junção entre as densidades [29], [30]. Será abordada aqui a primeira técnica desenvolvida, ou seja, aquela que faz uso do grafo completo. É apresentado, também, um algoritmo de treinamento para esta versão do OPF.

3.5.2 Classificador por Floresta de Caminhos Ótimos

A técnica de classificação supervisionada baseada em florestas de caminhos ótimos apresentada aqui modela as amostras como sendo os nós de um grafo completo. Os elementos mais representativos de cada classe do conjunto de treinamento, isto é, os protótipos, são escolhidos como sendo os elementos pertencentes às regiões de fronteiras entre as classes. Os protótipos participam de um processo de competição disputando as outras amostras oferecendo-lhes caminhos de menor custo e seus respectivos rótulos (ou classes). Ao final deste processo, obtêm-se um conjunto de treinamento particionado em árvores de caminhos ótimos, sendo que a união das mesmas remete a uma floresta de caminhos ótimos. Esta abordagem apresenta vários benefícios com relação a outros métodos de classificação de padrões supervisionados:

- É livre de parâmetros;
- Possui tratamento nativo de problemas de multiclases; e
- Não faz alusão sobre a forma e/ou separabilidade das classes.

O algoritmo OPF com grafo completo foi primeiramente apresentado por Papa et al. [28], [31] e tem sido amplamente utilizado em diversas aplicações para reconhecimento ou classificação de padrões. As próximas seções discutem a fundamentação teórica e os algoritmos de treinamento e classificação do algoritmo baseado em OPF utilizando grafo completo.

3.5.3 Fundamentação teórica

Seja Z uma base de dados λ -rotulada e Z_I e Z_3 os conjuntos de treinamento e teste, respectivamente, com $|Z_I|$ e $|Z_3|$ amostras, tais que $Z = Z_I \cup Z_3$. Seja $\lambda(s)$ uma função que associa o rótulo correto i , $i = 1, 2, \dots, c$ da classe i a qualquer amostra $s \in Z_I \cup Z_3$.

Seja $S \in Z_I$ um conjunto de protótipos de todas as classes (isto é, amostras que melhor representam as classes). Seja $\vec{v}$ um algoritmo que extrai n atributos de qualquer amostra $s \in Z_I \cup Z_3$, e retorna um vetor de atributos $\vec{v}(s) \in \mathfrak{R}^n$. Uma distância $d(s, t)$ entre duas amostras, s e t , é dada pela distância entre seus vetores de atributos $\vec{v}(s)$ e $\vec{v}(t)$. O problema consiste em usar S , (v, d) e Z_I para projetar um classificador ótimo, o qual pode prever o rótulo correto $\lambda(s)$ de qualquer amostra $s \in Z_3$. Assim sendo, foi proposto um classificador que cria uma partição discreta ótima, a qual é uma floresta de caminhos ótimos computada em $\mathfrak{R}^n$ pelo algoritmo da Transformada Imagem Floresta (IFT) [32].

Seja (Z_I, A) um grafo completo cujos nós são as amostras em Z_I , onde qualquer par de amostras define um arco em A (isto é, $A = Z_I \times Z_I$) (Figura 3.6a). Pode-se notar que os arcos não precisam ser armazenados e o grafo não precisa ser explicitamente representado.

Tem-se, também, que um caminho em um grafo é uma seqüência de amostras $\pi_{sk} = \langle s_1, s_2, \dots, s_k \rangle$, onde $(s_i, s_{i+1}) \in A$ para $1 < i < k - 1$. Um caminho é dito ser trivial se $\pi_s = \langle s \rangle$. Associa-se a cada caminho π_s um valor dado por uma função de valor de caminho f , denotada $f(\pi_s)$. Pode-se dizer que um caminho π_s é ótimo se $f(\pi_s) \leq f(\tau_s)$ para qualquer caminho τ_s , onde π_s e τ_s terminam numa mesma amostra s , independente de sua origem. Também denotamos $\pi_s \cdot \langle s, t \rangle$ a concatenação do caminho π_s com término em s e o arco (s, t) . O algoritmo da OPF pode ser utilizado com qualquer função de valor de caminho suave [32].

Uma função de valor de caminho f é suave quando, para qualquer amostra t , existe um caminho ótimo π_t o qual é trivial ou possui a forma $\pi_s \cdot \langle s, t \rangle$, onde

- $f(\pi_s) \leq f(\pi_t)$;
- π_s é ótimo; e
- Para qualquer caminho ótimo $\tau_s, f(\tau_s \cdot \langle s, t \rangle) = f(\pi_t)$.

Na versão OPF com grafo completo a função de custo abordada foi a f_{max} apresentada pela Equação (3.42), a qual é definida como:

$$f_{max}(\langle s \rangle) = \begin{cases} 0 & \text{se } s \in S, \\ +\infty & \text{caso contrário} \end{cases}$$

$$f_{max}(\pi_s \cdot \langle s, t \rangle) = \max\{f_{max}(\pi_s), d(s, t)\}, \quad (3.42)$$

Onde $f_{max}(\pi_s)$ computa a distância máxima entre amostras adjacentes em π_s , quando π_s não é um caminho trivial.

O algoritmo baseado em OPF associa um caminho ótimo $P^*(s)$ de S a toda amostra $s \in Z_I$, formando uma floresta de caminhos ótimos P (uma função sem ciclos, a qual associa a todo $s \in Z_I$ seu predecessor $P(s)$ em $P^*(s)$, ou uma marca *nil* quando $s \in S$, como mostrado na Figura 3.6d. Seja $R(s) \in S$ a raiz de $P^*(s)$ a qual pode ser alcançada por $P(s)$. O algoritmo calcula, para cada $s \in Z_I$, o custo $V(s)$ de $P^*(s)$, o rótulo $L(s) = \lambda(R(s))$ e o seu predecessor $P(s)$, como segue.

Algoritmo 1 – Classificador Supervisionado baseado em Floresta de Caminhos Ótimos usando grafo completo

- Entrada: Um conjunto de treinamento Z_I λ -rotulado, protótipos $S \subset Z_I$ e o par (v, d) para vetor de atributos e calculo das distancias;
- Saída: Floresta de caminhos ótimos P , mapa de valores de custo de caminhos V e mapa de rótulos L ; e
- Estruturas auxiliares: Fila de prioridades Q , e variável cst .

1. **Para todo** $s \in Z_I$, **Faça** $P(s) \leftarrow \text{nil}$ e $V(s) \leftarrow +\infty$.
2. **Para todo** $s \in S$, **Faça** $V(s) \leftarrow 0$, $P(s) \leftarrow \text{nil}$, $L(s) = \lambda(s)$ e **insira** s em Q .
3. **Enquanto** Q *é não vazia*, **Faça**
4. *Remova de* Q *uma amostra* s *tal que* $V(s)$ *é mínimo*.
5. **Para cada** $t \in Z_I$ *tal que* $s \neq t$ e $V(t) > V(s)$, **Faça**
6. *Calcule* $cst \leftarrow \max \{V(s), d(s, t)\}$.
7. **Se** $cst < V(t)$, **Então**
8. **Se** $V(t) = +\infty$, **Então** *remova* t *de* Q .
9. $P(t) \leftarrow s$, $L(t) \leftarrow L(s)$ e $V(t) \leftarrow cst$.
10. *Insira* t *em* Q .

As linhas 1–2 inicializam os mapas e inserem protótipos em Q . O laço principal calcula um caminho ótimo de S para cada amostra $s \in Z_I$ em uma ordem não decrescente de custos (linhas 3–10). A cada iteração, um caminho de custo ótimo $V(s)$ é obtido em P (linha 4). Empates são resolvidos em Q utilizando a política FIFO (*first-in-first-out*), ou seja, quando dois caminhos atingem uma determinada amostra s com o mesmo custo mínimo, s é associado ao primeiro caminho que o atingiu. O restante das linhas avalia se o caminho que atinge uma amostra adjacente t através de s é mais barato que o caminho que termina em t . Caso positivo, atualiza Q , $P(t)$, $L(t)$ e $V(t)$. No final do algoritmo, V armazena o valor do custo do caminho ótimo de S a cada amostra $s \in Z_I$ de acordo com f_{max} .

O algoritmo OPF com grafo completo utilizando f_{max} pode ser entendido como uma Transformada de Watershed usando a IFT [33] computada no espaço de atributos n -dimensional. Além dessa extensão, outra contribuição significativa é o processo de treinamento, o qual encontra protótipos ótimos (marcadores) na fronteira entre as classes afastando outliers do conjunto de treinamento, aumentando a exatidão do classificador.

3.5.4 Treinamento

A fase de treinamento do classificador baseado em floresta de caminhos ótimos usando o grafo completo consiste, basicamente, em encontrar o conjunto S de protótipos, ou seja, os elementos mais representativos de cada classe. Várias heurísticas poderiam ser

adotadas como, por exemplo, uma escolha aleatória de protótipos. Entretanto, tal escolha pode prejudicar o desempenho do classificador, tornando-o instável e com um alto grau de sensibilidade com relação aos protótipos escolhidos. Deseja-se, assim, estimar protótipos nas regiões de sobreposição de amostras e nas fronteiras entre as classes, visto que são regiões muito susceptíveis a erros de classificação.

Computando uma árvore geradora mínima (*Minimum Spanning Tree* - MST) no grafo completo (Z_I, A) , obtém-se um grafo conexo acíclico cujos nós são todas as amostras em Z_I , e os arcos são não direcionados e ponderados (Figura 3.6b). Seus pesos são dados pela distância d entre os vetores de atributos de amostras adjacentes. Esta árvore de espalhamento é ótima no sentido em que a soma dos pesos de seus arcos é mínima se comparada a outras árvores de espalhamento no grafo completo.

Os protótipos a serem escolhidos são os elementos conectados na MST com diferentes rótulos em Z_I , isto é, elementos mais próximos de classes diferentes (Figura 3.6c). Removendo-se os arcos entre classes diferentes, tais amostras adjacentes tornam-se protótipos em S e o Algoritmo 1 pode computar uma floresta de caminhos ótimos em Z_I (Figura 3.6d). Pode-se notar que uma dada classe pode ser representada por múltiplos protótipos (isto é, árvores de caminhos ótimos) e deve existir pelo menos um protótipo por classe.

O Algoritmo 1 pode computar uma floresta de caminhos ótimos com erro zero de classificação em Z_I , desde que a função $f_{\max}$ seja modificada. A idéia consiste, basicamente, em ponderar os arcos entre amostras de diferentes classes com um valor muito alto, impossibilitando assim que protótipos de uma classe conquistem elementos de outras classes. Desta forma, a função de valor de caminho $f_{\max}$ poderia ser escrita da seguinte forma:

$$f_{\max}(\langle t \rangle) = \begin{cases} 0 & \text{se } s \in S, \\ +\infty & \text{caso contrário} \end{cases} \quad (3.43)$$

$$f_{\max}(\pi_s \cdot \langle s, t \rangle) = \begin{cases} +\infty & \text{se } \lambda(t) \neq \lambda(s) \\ \max\{f_{\max}(\pi_s), d(s, t)\} & \text{caso contrário.} \end{cases}$$

A Equação (3.43) pondera todos os arcos $(s, t) \in A_k$ tais que $\lambda(t) \neq \lambda(s)$ com $d(s, t) = +\infty$, impedindo que tais arcos pertençam a algum caminho ótimo. Entretanto, vale destacar que, embora a Equação (3.43) garanta erro zero de classificação no conjunto de treinamento, a mesma produz resultados similares aos obtidos usando a Equação (3.42).

3.5.5 Classificação

Para qualquer amostra $t \in Z_3$, considera-se todos os arcos conectando t com amostras $s \in Z_1$, tornando t como se fosse parte do grafo original (ver Figura 3.6e, onde a amostra t é representada pelo triângulo no grafo). Considerando todos os possíveis caminhos entre S e t , deseja-se encontrar o caminho ótimo $P^*(t)$ de S até t com a classe $\lambda(R(t))$ de seu protótipo $R(t) \in S$ mais fortemente conexo. Este caminho pode ser identificado incrementalmente, avaliando o valor do custo ótimo $V(t)$ como:

$$V(t) = \min\{\max\{V(s), d(s, t)\}\}, \quad \forall s \in Z_1 \quad (3.44)$$

Seja $s^* \in Z_1$ o nó que satisfaz a equação acima (isto é, o predecessor $P(t)$ no caminho ótimo $P^*(t)$). Dado que $L(s^*) = \lambda(R(t))$, a classificação simplesmente associa $L(s^*)$ como a classe de t (Figura 3.6f). Um erro ocorre quando $L(s^*) \neq \lambda(t)$. Um procedimento similar é aplicado para amostras em um conjunto de avaliação Z_2 . Neste caso, entretanto, o uso das amostras em Z_2 tem como o intuito de aprender a distribuição das classes no espaço de atributos e, conseqüentemente, melhorar o desempenho de classificadores baseados em OPF.

Vale ressaltar que, embora a amostra a ser classificada esteja mais próxima de um elemento da classe bola (Figura 3.6e), a mesma é classificada como sendo da classe quadrado, o que demonstra que os classificadores baseados em OPF utilizam a força de conexidade entre as amostras para a classificação dos dados, ou seja, não são algoritmos baseados em conexidade local apenas como, por exemplo, os classificadores dos vizinhos mais próximos (*Nearest Neighbor* - NN) e k -vizinhos mais próximos (k -NN).

Ressalta-se também que o classificador baseado em OPF com grafo completo se assemelha ao algoritmo NN somente no caso onde todos os elementos do conjunto de treinamento são considerados protótipos, sendo este um caso atípico onde, certamente, o conjunto de atributos escolhido não foi o mais adequado para a representação do conjunto de dados. Outra diferença a ser considerada é que os classificadores NN e k -NN tomam uma decisão local para a classificação dos dados, ao contrário dos classificadores baseados em OPF, os quais possibilitam uma solução em âmbito global, criando uma floresta de caminhos

ótimos que mapeia, para cada amostra do conjunto de dados, o caminho ótimo entre ela e o seu protótipo mais fortemente conexo.

Figura 3.6. (a) Grafo completo. (b) MST do grafo completo. (c) Protótipos escolhidos. (d) Floresta de caminhos ótimos resultante. (e) Uma amostra de teste (triângulo). (f) O caminho ótimo do protótipo.

3.6 K-Vizinhos Mais Próximos

Apesar de não ser implementado nesta pesquisa, com o objetivo de completar a fundamentação teórica com relação as técnicas inteligentes, é apresentado o classificador K-Vizinhos Mais Próximos (k-NN, do inglês *K-Nearest Neighbor*), que é um método bastante utilizado em aplicações que envolvem tarefas de classificação, pois é de fácil entendimento e implementação. O k-NN é um classificador onde o aprendizado é baseado na analogia [34], [35].

A regra de classificação do k-NN [36], [37] é um método de classificação supervisionado e não-paramétrico, onde um padrão é dito pertencer a uma classe de acordo com a quantidade de vizinhos que pertençam a essa classe, conforme um critério de distância,

normalmente usa-se a distância Euclidiana, porém existem outras distâncias como a Manhattan e a Minkowski.

Seja $p=(p_1,p_2,...,p_n)$ e $q=(q_1, q_2,...,q_n)$ dois pontos de $\mathfrak{R}^n$:

- A distância Euclidiana entre p e q é dada por:

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (3.45)$$

- A distância Manhattan entre p e q é dada por:

$$d(p,q) = |p_1 - q_1| + |p_2 - q_2| + \dots + |p_n - q_n| \quad (3.46)$$

- A distância Minkowski entre p e q é dada por:

$$d(p,q) = \left(|p_1 - q_1|^j + |p_2 - q_2|^j + \dots + |p_n - q_n|^j \right)^{1/j} \quad (3.47)$$

Onde $j \in \mathfrak{R}$.

A distância Minkowski é a generalização das duas distâncias anteriores. Quando $j = 1$, esta distância representa a distância de Manhattan e quando $j = 2$, a distância Euclidiana.

Os classificadores representam diferentes paradigmas, sendo que o k-NN faz parte do chamado *lazy learning*, pois não há um treinamento explícito, considerando que cada padrão de treinamento é adicionado nos vetores de treinamento, diferente do SVM e da ANN-MLP com *backpropagation* que estão, respectivamente, no contexto de aprendizado estatístico e conexionista. Logo, uma de suas principais vantagens do k-NN é o pequeno tempo de treinamento, pois se resume apenas a adicionar padrões nos vetores. O conjunto de treinamento é formado por vetores n-dimensionais e cada elemento deste conjunto representa um ponto no espaço n-dimensional.

O método k-NN é um método de classificação que não possui processamento na fase de treinamento. Para cada padrão de teste é medida a distância entre ele e todos os padrões de treinamento. Verifica-se a quais classes pertencem os k padrões mais próximos e a classificação é feita, associando-se o padrão de teste à classe que for predominante [35].

Algoritmo 2 – Classificador k -NN

1. **Para** cada padrão de treinamento p de entrada q , **faça**:
2. *Calcular distância entre p e q .*
3. *Selecionar os k padrões mais próximos.*
4. *Determinar a classe do novo padrão q , associando com a classe de maior ocorrência entre os k padrões, em caso de distâncias iguais entre padrões de classes diferentes, determine a nova classe aleatoriamente.*

A classificação pode ser para o caso em que o parâmetro k é igual a 1 (NN), e se o k for maior que um, por exemplo, $k=3$, na qual são considerados três vizinhos do novo padrão, ou seja, são analisadas as três menores distâncias do novo padrão para os padrões de treinamento. A classe que tiver o maior número de padrões entre essas distâncias é que irá determinar a classe do novo padrão. A Figura 3.7 ilustra esse processo de classificação. Para classificar um ponto, primeiro tomam-se os k -vizinhos mais próximos dele e, dentro desse conjunto, encontra-se a classe mais representativa. Na Figura 3.7, o ponto desconhecido 1 será classificado como classe B e o ponto desconhecido 2 será classificado como classe A.

Figura 3.7. Processo de classificação do k -NN.

Na escolha do parâmetro k deve-se ter cuidado para escolher um número ímpar, pois um número par poderia causar conflito, quando um ponto tivesse o mesmo número de vizinhos de cada classe, por exemplo.

A classificação k-NN é bem simples, porém tem a desvantagem de ter que armazenar todos os dados na memória, ou seja, gerar classificadores complexos, por causa do armazenamento, além da necessidade de se realizar muitos cálculos de distância.

O k-NN é um classificador que possui apenas um parâmetro livre (o número de k-vizinhos) que é controlado pelo usuário com o objetivo de obter uma melhor classificação.

Este processo de classificação pode ser computacionalmente exaustivo se considerado um conjunto com muitos dados. Para determinadas aplicações, no entanto, o processo é bem aceitável.

3.7 Algoritmos Genéticos

Apesar de também não ser utilizada neste trabalho, é importante conhecer a técnica dos Algoritmos Genéticos (GA), que são algoritmos de busca e otimização baseados na analogia com os processos de seleção natural e genética evolucionária. A essência do método consiste em manter uma população de indivíduos (cromossomos), os quais representam possíveis soluções para um problema específico. A melhor solução é atingida através de um processo de seleção, competitivo, envolvendo cruzamentos e mutações [38].

O que torna um GA atrativo é sua habilidade para acumular informações sobre um espaço de busca inicial desconhecido e explorar este conhecimento para levá-lo a buscas subseqüentes em subespaços úteis. Desde a formulação inicial, em 1975, os GA vêm sendo empregados para solucionar problemas complexos, cujas propriedades das funções objetivo não são boas, tal como continuidade e diferenciabilidade [38].

4 ABORDAGENS INTELIGENTES PARA IDENTIFICAÇÃO DE CONSUMIDORES COM INSTALAÇÕES IRREGULARES

Neste capítulo serão mencionados trabalhos baseados em métodos que usam os sistemas inteligentes e métodos estatísticos para classificar e identificar perfis de consumidores com instalações irregulares.

Eller [6], em sua tese de doutorado, propõe um método para auxiliar no gerenciamento de perdas comerciais de energia elétrica através da mineração de dados, destacando-se o uso das redes neurais para descobrir comportamentos suspeitos de consumidores de energia, indicando a possibilidade de fraude ou não.

O trabalho proposto em [39] apresenta a capacidade de técnicas computacionais modernas como ferramentas na classificação de consumidores de energia elétrica e a extração de padrões de comportamento destes através de seus históricos de consumo, representados pelas curvas de carga. A classificação é feita através do uso de redes neurais, utilizando métodos de análise dos dados das curvas de carga. As curvas de carga diárias são obtidas através de medições reais utilizando medidores eletrônicos e o mapa auto-organizável de Kohonen [15] é utilizado como ferramenta de classificação. Segundo o artigo [39], existem inúmeras técnicas e ferramentas de estudo de comportamento. Seu objetivo principal é agrupar consumidores com curvas de carga semelhantes.

Masselli [40] apresenta um modelo baseado no hibridismo entre redes neurais artificiais e algoritmos genéticos, capaz de identificar na curva de carga adquirida, em um ponto qualquer do sistema de distribuição, as porções de consumo relativas a cada um dos principais setores consumidores presentes naquele ponto. Seus resultados iniciais indicaram um melhor desempenho da rede neural convencional, porém, à medida que se aumentou a complexidade dos padrões de treinamento, o sistema híbrido passou a apresentar resultados cada vez melhores.

Paula [41] propôs a caracterização de consumidores por suas curvas de carga, aplicando Redes Neurais na análise dos dados. As curvas de carga foram estabelecidas com base nas medições de consumidores de baixa e média tensão, utilizando equipamentos de medição digitais instalados por uma concessionária. O conjunto de medições foi submetido a uma avaliação em busca de inconsistências na massa de dados, através de uma etapa de limpeza e pré-processamento. Em seguida, foi feito um processo de redução da massa de dados, onde cada consumidor passa a ser representado por uma única curva representativa

composta pela média das curvas medidas para este consumidor. Para realizar devidamente a comparação de padrões para o agrupamento, as curvas dos consumidores devem ser trazidas para uma mesma escala. Esta normalização pode ser obtida utilizando o valor da demanda média como valor de base. O conjunto de dados foi, então, dividido em grupos menores de acordo com a finalidade da utilização da energia elétrica, o nível da tensão, o dia da semana e a grandeza medida (potência ativa e reativa). Em seguida, são utilizadas Redes Neurais auto-organizadas para a formação das curvas típicas de carga para cada classe de consumidor. A definição do número ideal de agrupamentos se deu através de funções de avaliação, buscando um número ótimo de grupos. O resultado desta análise foi a formação de curvas típicas representativas para permitir uma avaliação de carga do mercado consumidor de baixa e média tensão da distribuidora.

Monederon et al. [42] também usou técnicas de mineração de dados, propondo a utilização Redes Neurais Artificiais - ANN [14], juntamente com uma análise estatística de fraudes no consumo elétrico.

Ferreira [43] afirma que o combate às fraudes de energia deve ser feito em tempo real do processo de venda da energia, visando minimizar ações contrárias aos interesses da empresa. Para isto, é apresentada uma metodologia que consiste em duas etapas: identificar os transformadores que apresentem considerável diferença entre a energia fornecida e a faturada e a identificação de consumidores com maior possibilidade de erro de medição, fraude ou desvio, dentre os atendidos por cada transformador selecionado na etapa anterior. A primeira etapa é feita com os dados da concessionária através do cálculo da perda comercial de cada transformador, feito pela subtração da energia faturada de todos os consumidores atendidos pelo transformador, mais as perdas técnicas, da energia demandada no transformador, medida por um Módulo de Aquisição de Dados (MAD). A segunda etapa é feita pela avaliação do histórico dos consumidores atendidos pelo transformador selecionado na primeira etapa. Esta avaliação ocorre com a criação de regras que serão aplicadas sobre os dados destes consumidores. Primeiramente, é comparado o consumo do mês em questão com a média dos consumos, em três regras: nos 3, 6 e 12 últimos meses. Em seguida, uma nova regra analisa a característica sazonal do consumidor a fim de se evitar que consumidores normais sejam classificados como fraudadores. Uma terceira regra avalia somente o consumo do mês de referência para identificar anormalidades em consumidores que estão fraudando energia a mais de um ano.

A metodologia proposta em [44] apresenta um sistema de identificação de possíveis fraudes em consumidores de alta tensão. Foram utilizados os dados de medição de uma

empresa brasileira, contendo aproximadamente 2000 consumidores. Estes dados são submetidos a uma rede neural chamada Mapas Auto-Organizáveis (SOM) [15], que identifica grupos de registros com características semelhantes (*clusters*). Os dados de cada consumidor foram separados por medições semanais, e, através da técnica SOM, foram agrupados em 2 clusters: o primeiro, com as curvas de carga típicas dos consumidores, e o segundo, com curvas atípicas, apresentando um consumo médio baixo. Este consumo baixo foi obtido pela redução proposital em 30% do consumo em um período determinado do dia. Então, os consumidores do cluster 1 foram classificados como normais, enquanto os consumidores do cluster 2 foram classificados como suspeitos.

O artigo de Nagi et al., apresenta um trabalho com uma nova abordagem para análise de perdas não-técnicas por concessionárias de energia usando SVM [45]. A principal motivação deste estudo é a de auxiliar a concessionária Tenaga Nasional Berhad (TNB) na Malásia para reduzir as perdas não-técnicas no setor da distribuição de eletricidade, devido ao roubo. O modelo proposto pré-seleciona consumidores suspeitos para serem inspecionados no local de fraude baseada nas irregularidades e no comportamento anormal de consumo. Esta abordagem proporciona um método de extração de dados e envolve características históricas de consumo a partir dos dados dos consumidores. A abordagem SVM utiliza as informações do perfil de carga dos consumidores para expor comportamento anormal que é conhecido por ser altamente correlacionada com as perdas não-técnicas. O resultado de classificação das classes que são utilizadas para direcionar os suspeitos de fraude, durante a inspeção no local, é baseado no comportamento suspeito devido às significativas irregularidades no consumo. Os resultados de simulações provam que o método proposto é mais eficaz em comparação com as atuais medidas tomadas por TNB, a fim de reduzir as perdas não-técnicas.

Em outro trabalho muito semelhante, também publicado por Nagi et al. [46], uma abordagem híbrida entre Algoritmos Genéticos - GA [47] e SVM foi também aplicada para detecção perdas não-técnicas. O GA proporciona uma maior convergência e uma globalização otimizada dos hiper-parâmetros do SVM utilizando uma combinação aleatória e de genomas pré-povoados.

O método proposto por Nizar [48] apresenta um estudo para a escolha do melhor método para a geração de curvas de carga. Neste método, são utilizadas técnicas de aglomeração (*clusters*) e mineração de dados. Após a coleta de dados de consumidores divididos em categorias, como residencial, comercial e industrial, estes dados são separados em clusters, cada um contendo um subconjunto de dados com características semelhantes. Em seguida, os dados do subconjunto são usados para classificar os consumidores de acordo com

o comportamento, sendo, então, criados perfis de consumo típicos. Estes perfis típicos são comparados com os perfis de consumo a serem averiguados. Se alguma destas novas curvas apresentarem desvio com relação à curva típica, esta é selecionada para investigação. Se for confirmado que esta variação se deve à ocorrência de perdas comerciais, a nova curva é atualizada e utilizada futuramente para fins de previsão.

Nizar et al. também usa a mineração de dados aplicadas em técnicas de análise perdas não-técnicas [49], [50]. O método é baseado em diversos algoritmos populares de classificação e envolve a extração dos padrões de comportamento do consumo de consumidores a partir de dados históricos e da organização dos dados em médias anuais, mensais, semanais e diárias. As diversas técnicas de classificação são utilizadas e comparadas. A intenção é garantir a obtenção de melhores resultados no desenvolvimento de perfis de carga representante para ser utilizado como referência para as análises de perdas não-técnicas direcionada a detectar quaisquer atividades significativas que possam contribuir para estas perdas.

Em [51] e em [52], Nizar et al. também apresenta uma nova abordagem para análise de perdas não-técnicas (NTL-*Nontechnical Losses*) de concessionárias usando as técnicas computacionais modernas com extrema aprendizagem de máquina (ELM - *Extreme Learning Machine*). Segundo Nizar et al., as perdas não-técnicas são significativas em diversos países. A abordagem ELM apresentada utiliza a informação de um perfil de carga de um consumidor para expor comportamento anormal que é conhecido por ser altamente correlacionada com as atividades de perdas não-técnicas. Esta abordagem proporciona um método de extração de dados para esta finalidade, e que envolve extrair padrões de comportamento dos clientes a partir de dados históricos de consumo. Os resultados das classificações das classes são usados para revelar qualquer comportamento significativo devido a irregularidades no consumo. Nestes artigos, os algoritmos ELM e o ELM de seqüenciamento online (OS-ELM) são utilizados para alcançar uma melhor classificação para aumentar o desempenho e a precisão dos resultados. Uma comparação desta abordagem de classificação com a técnica de máquinas de vetores de suporte (SVM) é também realizada, e os resultados com ELM para a precisão na análise NTL demonstraram serem superiores.

O trabalho realizado por Ortega [9] desenvolveu uma metodologia que identifica, com maior precisão, o perfil do cliente irregular (comprovada fraude no medidor, furto por ligação clandestina ou irregularidade técnica). O sistema inteligente resultante, denominado SIIPERCOM, baseia-se em Redes Neurais, para o agrupamento dos clientes com

comportamentos semelhantes e para a classificação dos clientes de cada grupo em ‘normais’ ou ‘irregulares’.

5 DIRETRIZ METODOLÓGICA DA PESQUISA

5.1 Introdução

A fim de cumprir com o objetivo desta pesquisa, que é desenvolver ferramentas computacionais inteligentes para detectar as perdas comerciais, através de informações presentes nas bases de dados das concessionárias, identificando possíveis inconsistências e incoerências nos dados de medição de forma automática e servindo de apoio à decisão a ser tomada, foram propostos métodos alternativos para explorar os modelos das técnicas computacionais com algoritmos inteligentes, através das Redes Neurais (ANN-MLP), Mapas Auto-Organização (SOM), Máquinas de Vetores De Suporte (SVM) e Floresta de Caminhos Ótimos (OPF).

5.2 Utilização da Base de Dados

Nesta seção é descrito quais são as informações da base de dados usadas nos experimentos, ou seja, as características extraídas de cada perfil de consumidor que são utilizadas nas arquiteturas dos algoritmos inteligentes num processo de mineração de dados (*Data Mining*).

Para o desenvolvimento desta pesquisa, uma concessionária de energia brasileira disponibilizou uma base de dados com perfis de consumidores comerciais e industriais, onde as fraudes são maiores e causam grandes impactos. Todas as informações dos consumidores foram inspecionadas por técnicos da concessionária dando autenticidade aos valores da base de dados.

Cada perfil de consumidor é representado com as seguintes informações [53]:

- Demanda Contratada: é o valor de demanda a ser obrigatória e continuamente disponibilizada pela concessionária, conforme valor e período de vigência do contrato. Este valor, que é contratado pelo consumidor, deverá ser pago à concessionária, seja ou não utilizada, expressa em quilowatts (kW);

- Demanda Medida (D_{max}): é a maior demanda de potência ativa, verificada por medição, no intervalo de quinze minutos durante o período de faturamento, expressa em quilowatts (kW);
- Potência Instalada (P_{inst}): soma das potências nominais de equipamentos elétricos instalados e em condições de entrar em funcionamento na unidade consumidora, expressa em quilowatts (kW); e
- Fator de Carga (FC): é a razão entre a demanda média (D_{media}) e a demanda máxima da unidade consumidora, registradas em um mesmo intervalo de tempo, logo:

$$FC = \frac{D_{media}}{D_{max}} \quad (5.1)$$

FC é um índice que permite verificar o quanto que a energia elétrica é utilizada de forma racional e D_{media} é definida como a razão entre a energia total (ε_T) e o período T , e é descrito por:

$$D_{media} = \frac{\varepsilon_T}{T} \quad (5.2)$$

E ε_T é dado por:

$$\varepsilon_T = \int_0^T D(t) \cdot dt \quad (5.3)$$

Em que $D(t)$ é a curva da demanda, como na Figura 5.1.

Figura 5.1. Gráfico de interpretação das informações de cada perfil de consumidor.

É interessante dar uma atenção especial ao fator de carga (FC), que é um índice que varia de 0 a 1 definido por (5.1). Deste modo, tem-se que quanto maior o valor do fator de carga obtido, melhor terá sido a utilização das cargas elétricas ao longo do tempo. Por outro lado, um fator de carga baixo pode indicar que houve concentração de consumo de energia elétrica em um curto período de tempo, determinando alterações na demanda.

A seguir, algumas situações que conduzem a esses valores baixos:

- Equipamentos de grande potência, operando a plena carga somente algumas horas do período de utilização, funcionando com carga reduzida ou sendo desligados nos demais períodos;
- Cargas de grande porte ligadas simultaneamente;
- Curtos-circuitos e fugas de corrente; e
- Falta de programação para utilização de energia.

A melhoria do fator de carga, além de diminuir as despesas com energia consumida, conduz a um melhor aproveitamento e aumento da vida útil de toda a instalação elétrica, inclusive de motores e equipamentos, e a uma otimização dos investimentos nas instalações.

O aumento do fator de carga pode ser conseguido através de medidas que, na sua maioria, não implica em investimentos. Estão relacionadas, a seguir, algumas delas:

- Selecionar e reprogramar os equipamentos e sistemas que possam operar fora do horário de maior demanda da instalação, fazendo um cronograma de utilização de seus equipamentos elétricos, anotando a capacidade e o regime de trabalho de cada um, através de seus horários de funcionamento;

- Evitar partidas simultâneas de motores que iniciem operação com carga;
- Diminuir, sempre que possível, a operação simultânea dos equipamentos; e
- Verificar se a manutenção e a proteção da instalação elétrica e dos equipamentos são adequadas, de modo a se evitar a ocorrência de curtos-circuitos e fugas de corrente.

5.2.1 Base de Dados do Primeiro Ensaio

A base de dados do primeiro ensaio (Bt) possui 736 consumidores industriais e está dividida em 116 consumidores irregulares (fraudador) e 620 consumidores normais (não fraudador).

Este primeiro ensaio possuía um caráter preliminar, ou seja, tratava-se de um teste para observar qual o rumo que o desenvolvimento desta pesquisa deveria tomar.

5.2.2 Base de Dados do Segundo Ensaio

Para o segundo ensaio foram utilizadas duas bases de dados, uma base (Bi) com 5190 perfis de consumidores industriais, dividida em 334 consumidores irregulares e 4856 consumidores normais, e outra base (Bc) com 8067 perfis de consumidores comerciais, sendo 473 irregulares e 7594 consumidores normais.

5.3 Arquitetura Inteligente

A arquitetura é uma maneira de se observar o que está sendo proposto nesta dissertação, considerando que a partir de dados de perfis de consumidores, pode-se desvendar a possibilidade de um consumidor estar fraudando ou não.

A Figura 5.2 ilustra uma forma geral da arquitetura proposta.

Figura 5.2. Esquema geral da arquitetura inteligente proposta.

5.4 Bibliotecas Utilizadas na Implementação dos Classificadores

Todos os classificadores inteligentes foram implementados em linguagem C/C++ através de pacotes e bibliotecas de códigos abertos na internet. Para o SVM-RBF foi usado a última versão do pacote LibSVM [27] com Função de Base Radial, com parâmetros de otimização e estratégia um-contra-um para resolver problemas de multiclasse. Para o SVM-LINEAR, foi usado o pacote LibLINEAR [54] com o parâmetro $C=1$. Para o OPF foi usado a biblioteca LibOPF [55] e para a ANN-MLP foi usando a biblioteca Fast Artificial Neural Network Library (FANN) [56]. A arquitetura da ANN-MLP foi escolhida empiricamente, pois não existe na literatura uma maneira de realizar a melhor escolha da arquitetura neural, de modo que se tenha o melhor desempenho, para se definir os números de camadas intermediárias e de neurônios da ANN-MLP.

6 RESULTADOS E DISCUSSÕES

6.1 Introdução

Entre as abordagens supervisionadas, ANN e SVM são as mais populares, e o OPF é uma abordagem recente e inovadora, logo experimentos são feitos para comparações entre os métodos inteligentes utilizados para a identificação de possíveis consumidores irregulares (fraudador).

O trabalho apresenta dois ensaios, sendo que em cada um deles foram feitas duas séries de experimento, na primeira série foram usadas 50% da base de dados para treinamento e os 50% restantes para teste dos classificadores a fim de calcular a média de exatidão com os seus respectivos desvios-padrão e dos tempos de treinamento e de execução dos testes em segundos. Na segunda série foram executados experimentos com diferentes porcentagens de tamanho dos conjuntos de treinamentos para fazer uma comparação da robustez dos classificadores de acordo com essa variação.

6.2 Primeiro Ensaio (Base Bt)

No primeiro ensaio, para ambas as séries de experimentos, foram implementados o OPF, SVM-RBF (SVM com função de base radial como a função núcleo), SVM-LINEAR (SVM com função núcleo linear) e a ANN-MLP (Perceptron multicamadas com treinamento através do algoritmo backpropagation) com 10 execuções para cada técnica com dados para conjuntos de treinamentos e testes gerados de forma aleatória para calcular a média de exatidão e dos tempos de treinamento e execução dos testes.

A configuração da ANN-MLP é: 4 atributos (características) na camada de entrada, duas camadas intermediárias com 92 neurônios cada, e uma camada de saída com duas classes distintas (se o consumidor é um possível fraudador ou não). Na tentativa de padronizar o número de neurônios das camadas intermediárias, devido ao empirismo da arquitetura neural, escolheu-se utilizar 92 neurônios em cada camada intermediária, na qual

representa 25% (ou $\frac{1}{4}$) do número de amostras do conjunto treinamento. A função de ativação sigmoideal foi usada nas camadas intermediárias e na camada de saída.

6.2.1 Avaliação dos Classificadores

Os classificadores OPF, SVM-RBF, SVM-LINEAR e a ANN-MLP foram avaliados para a identificação de perdas comerciais entre consumidores usando 50% dos dados para treinamento e 50% para teste. A Tabela 6.1 apresenta as médias de exatidão e dos tempos de treinamento e classificação após 10 execuções com conjuntos de treinamentos e testes gerados aleatoriamente.

Tabela 6.1. Médias de Exatidão e Médias de Tempo de Treinamento e de Classificação pelo OPF, SVM-RBF, SVM-LINEAR e ANN-MLP para a Base Bt

Classificador	Exatidão	Tempo de treinamento (segundos)	Tempo de classificação (segundos)
OPF	90,21 $\pm$ 2,93	0,0257	0,0223
SVM-RBF	88,93 $\pm$ 3,07	13,4817	0,0222
SVM-LINEAR	45,40 $\pm$ 6,31	2,4514	0,0048
ANN-MLP	53,01 $\pm$ 6,95	1708,8500	0,0078

Pode-se notar que o OPF e o SVM-RBF tiveram resultados similares e tiveram um desempenho superior que o SVM-LINEAR e a ANN-MLP. O OPF foi 524,57 vezes mais rápido do que o SVM-RBF na fase de treinamento. Deste modo, o OPF é capaz de realizar treinamentos em tempo real, considerando a possibilidade de que novas amostras possam ser adicionadas ao conjunto de treinamento, necessitando de novos treinamentos para a atualização do classificador. Similarmente, pode-se ter um sistema de detecção em tempo real para identificação das perdas comerciais sendo que novos perfis de consumidores podem ser adicionados e o sistema ser re-treinado o mais rápido possível, diminuindo os custos para as empresas de energia.

6.2.2 Robustez dos Classificadores

Nesta seção é avaliada a robustez dos classificadores de acordo com a variação no tamanho do conjunto treinamento. Foram repetidos os experimentos realizados na seção anterior com diferentes porcentagens de tamanho dos conjuntos de treinamentos e testes para o OPF, SVM-RBF, SVM-LINEAR e para a ANN-MLP. A Figura 6.1 apresenta as médias de exatidão sobre os testes depois de 10 execuções para cada porcentagem do tamanho dos conjuntos de treinamento e de teste.

Figura 6.1. Robustez dos classificadores a respeito de diferentes porcentagens de tamanho do conjunto de treinamento e teste (Bt).

Pode-se observar que apenas o OPF e o SVM-RBF apresentaram desempenhos similares, aumentando suas médias de exatidão à medida que se aumentam as porcentagens de treinamento. A ANN-MLP e o SVM-LINEAR também tiveram desempenhos semelhantes entre si, mas com baixas taxas de reconhecimento de padrões. Isto reflete como os tradicionais métodos de classificação (ANN-MLP, por exemplo) falham no tratamento de problemas não separáveis. A escolha empírica de uma arquitetura de rede neural mais adequada torna-se inviável para grandes bases de dados, como é o caso da utilizada no presente trabalho.

A ANN-MLP, também, é um modelo instável (o desempenho oscila muito dependendo do conjunto de treinamento) e seu desempenho pode melhorar quando combinamos várias redes [57], gerando um maior custo computacional. Mesmo assim, o aumento no número de redes pode também provocar piora de exatidão e o número ideal de redes é desconhecido [58]. Como a premissa de separabilidade no espaço de atributos não é sempre válida, SVM busca uma separabilidade linear ótima em um espaço de maior dimensão. Como desvantagens, esta premissa pode não ser válida com dimensão finita [59] e o custo computacional do SVM aumenta muito rápido com o número de vetores de suporte [60].

Estas situações fazem do OPF um bom classificador, pois é livre de parâmetros e não faz alusão quanto à forma e a separabilidade no espaço de atributos, e tem uma fase de treinamento mais rápida.

6.3 Segundo Ensaio (Bases Bi e Bc)

No segundo ensaio repetiu-se a mesma metodologia do primeiro ensaio, já que os primeiros resultados obtidos foram bastante satisfatórios, mas com o acréscimo do classificador SOM e com a aplicação de bases de dados maiores de diferentes perfis de consumidores (industriais e comerciais) de maneira a aperfeiçoar e aprimorar ainda mais os resultados alcançados até o momento na identificação de perdas comerciais.

A configuração da ANN-MLP é: 4 atributos (características) na camada de entrada, apenas uma camada intermediária com 10 neurônios, e uma camada de saída com duas

classes distintas (se o consumidor é fraudador ou não). A função de ativação sigmoidal foi usada na camada intermediária e na camada de saída.

Algumas mudanças no número de neurônios e no número de camadas intermediárias da ANN-MLP foram feitas após a realização de alguns testes, pois ao analisar os seus resultados, constatou-se que não havia necessidade de se fazer um ensaio com duas camadas intermediárias e com grande quantidade de neurônios, como feito no primeiro ensaio, ou seja, percebeu-se que a padronização, em 25% das amostras do conjunto treinamento para a definição do número de neurônios de cada camada intermediária, tornaria o tempo para a execução do treinamento da ANN-MLP muito grande, e isto seria inviável, quando se comparado com as demais técnicas inteligentes.

6.3.1 Avaliação dos Classificadores

Os classificadores OPF, SVM-RBF, SVM-LINEAR, ANN-MLP e o SOM foram avaliados para a identificação de perdas comerciais entre consumidores usando 50% dos dados para treinamento e 50% para teste. As Tabelas 6.2 e 6.3 apresentam as médias de exatidão e dos tempos de treinamento e classificação após de 10 execuções com conjuntos de treinamentos e testes gerados aleatoriamente.

Tabela 6.2. Médias de Exatidão e Médias de Tempo de Treinamento e de Classificação pelo OPF, SVM-RBF, SVM-LINEAR, ANN-MLP e SOM para a Base Bi

Classificador	Exatidão	Tempo de treinamento (segundos)	Tempo de classificação (segundos)
OPF	86,62 ± 2,28	1,0343	0,9102
SVM-RBF	78,72 ± 3,72	890,6785	0,3622
SOM	59,56 ± 2,72	259,7321	0,3070
SVM-LINEAR	50,52 ± 2,34	175,7613	0,0073
ANN-MLP	50,26 ± 0,78	159,9900	0,0007

Tabela 6.3. Médias de Exatidão e Médias de Tempo de Treinamento e de Classificação pelo OPF, SVM-RBF, SVM-LINEAR, ANN-MLP e SOM para a Base Bc

Classificador	Exatidão	Tempo de treinamento (segundos)	Tempo de classificação (segundos)
OPF	88,29 ± 1,48	1,5836	1,3918
SVM-RBF	81,10 ± 2,57	1500,6685	0,4960
SOM	62,92 ± 5,03	341,7424	0,4059
SVM-LINEAR	50,42 ± 0,89	237,0378	0,0086
ANN-MLP	50,99 ± 1,41	279,8700	0,0014

Pode-se observar, através das Tabelas 6.2 e 6.3, que o OPF superou todos os classificadores, tanto quanto na exatidão quanto na soma do tempo de treinamento e classificação dos consumidores. É muito importante que as concessionárias de energia possam diagnosticar as perdas comerciais o quanto antes, ou seja, em tempo real, a fim de evitar maiores prejuízos.

6.3.2 Robustez dos Classificadores

Nesta seção é avaliada a robustez dos classificadores de acordo com a variação no tamanho do conjunto treinamento. Foram repetidos os experimentos realizados na seção anterior com diferentes porcentagens de tamanho dos conjuntos de treinamentos e testes para o OPF, SOM, SVM-RBF, SVM-LINEAR e para a ANN-MLP. As Figuras 6.2 e 6.3 apresentam as médias de exatidão sobre os testes depois de 10 execuções para cada porcentagem do tamanho dos conjuntos de treinamento e de teste.

Figura 6.2. Robustez dos classificadores a respeito de diferentes porcentagens de conjunto de treinamento e teste (Bi).

Figura 6.3. Robustez dos classificadores a respeito de diferentes porcentagens de conjunto de treinamento e teste (Bc).

Através das Figuras 6.2 e 6.3, pode-se notar a superioridade do OPF quanto à utilização de diferentes porcentagens de tamanho dos conjuntos de treinamento e teste. Logo, o classificador OPF é mais robusto que os demais classificadores, demonstrando ser bastante eficiente entre os conjuntos de treinamento avaliados.

6.4 Software de Aplicação Desenvolvido

Neste trabalho foi desenvolvido um software para possibilitar uma interação amigável na aplicação dos algoritmos inteligentes na identificação de possíveis fraudes.

O software é composto de apenas uma interface gráfica que foi desenvolvida no ambiente do MATLAB e é apresentada na Figura 6.4.

Figura 6.4. Interface gráfica para o software desenvolvido.

Observando a Figura 6.4, no campo onde são selecionados os dados do treinamento, o usuário define a técnica inteligente que será aplicada, e o software carrega toda a

implementação do algoritmo inteligente desenvolvido para classificar os perfis de consumidores.

Após a definição do treinamento da técnica inteligente escolhida, seleciona-se o banco de dados (ou base de dados) para que seja realizada a classificação dos consumidores.

É necessário que a base de dados esteja em uma planilha no formato padrão do Excel do Microsoft Office, ou seja, a extensão do arquivo deve ser “xls”. Nessa planilha, algumas regras que deverão ser cumpridas para que a simulação seja realizada corretamente, caso contrário, alguns erros impedirão que os consumidores sejam classificados. As regras para a formatação da planilha são apresentadas na Tabela 6.4.

Tabela 6.4. Regras para a Formatação da Planilha com os Dados de Cada Perfil de Consumidores no Padrão do Excel

Coluna	Dados de Informações dos Consumidores
1	Demanda Contratada
2	Demanda Medida
3	Fator de Carga
4	Potência Instalada

Como se pode observar a Tabela 6.4, os dados da planilha são as mesmas características, apresentadas na Seção 5.2, que serviram para o treinamento dos sistemas inteligentes. Deste modo, o software consegue realizar a classificação dos consumidores.

Após a seleção dos dados de treinamento e da base de dados a ser classificada, o software realiza uma simulação, integrando todas as informações disponibilizadas pelo usuário com os classificadores inteligentes.

Observando a Figura 6.5, os resultados da simulação para a classificação dos consumidores pelo software desenvolvido são apresentados da seguinte maneira:

- É apresentado o total de possíveis consumidores fraudulentos;
- São apresentados quais são esses possíveis fraudulentos, através de uma lista.
- É apresentado o total de consumidores normais (ou regulares); e
- É apresentado um gráfico, para melhor visualização, com todos os consumidores.

Figura 6.5. Resultado da classificação dos consumidores feita pelo software desenvolvido.

Para a simulação apresentada na Figura 6.5, foi realizada uma simulação-teste com 110 perfis de consumidores, sendo que 17 deles são reais fraudadores, utilizando a aplicação de um classificador ANN-MLP com uma camada intermediária, contendo 10 neurônios, e função de ativação sigmoideal. Após a simulação realizada pelo software, são apontados 20 consumidores que podem ser possíveis fraudadores, ou seja, dentre os que realmente são fraudadores, o algoritmo inteligente treinado só deixou de identificar 1 (um) consumidor que comete fraude. Embora tenha detectado 4 consumidores normais como sendo possíveis fraudadores, o resultado é bastante otimista, pois todos os fraudadores foram identificados, o que comprova a aplicabilidade do software. Assim, a aplicação desse software pode ser útil para uma concessionária de energia, pois diminuiria o número de inspeções de campo feitas por técnicos para comprovar a fraude, evitando o desperdício de recursos.

Note que apesar dos resultados com a ANN-MLP não tenham sido satisfatórios nas Seções 6.2 e 6.3, para esta simulação-teste com o software, obteve-se um ótimo desempenho, que pode ser justificado pela boa generalização das características escolhidas, ou seja, existe um balanceamento entre as amostras das classes “fraudadores” e “não fraudadores” no conjunto treinamento, nas quais estão mais representativas, mas ainda é recomendada a utilização do classificador OPF para esta tarefa de classificação.

6.5 Observações

Os trabalhos apresentados no Capítulo 4 possuem algumas diferenças quanto à metodologia desta pesquisa, sendo que alguns fazem, primeiramente, uma estimação da curva de carga dos perfis de consumidores para depois fazer a análise de uma possível irregularidade, outros usam outros tipos de variáveis de entrada da arquitetura inteligente e até mesmo outros tipos de classificadores. Os resultados apresentados nas Tabelas 6.1, 6.2 e 6.3 com os classificadores SVM e ANN foram próximos dos resultados de exatidão encontrados na literatura. A utilização do classificador OPF neste trabalho de pesquisa se destaca por apresentar uma exatidão mais alta quando comparado aos resultados dos outros classificadores mais tradicionais já utilizados em pesquisas internacionais relacionadas com as perdas comerciais [45], [46], [51], [52].

7 CONCLUSÕES

7.1 Conclusões Gerais

Os investimentos realizados nas concessionárias têm como principal objetivo uma melhora significativa de seu desempenho financeiro e técnico, buscando, maior produtividade, eficiência e rentabilidade. Com a finalidade de se obter uma melhor gestão com relação às perdas de energia, uma das formas de maximizar a energia disponível para comercialização é a concessionária de energia promover a diminuição de fraudes de energia.

Embora a literatura estudada apresente avanços nesta área, pesquisas de métodos alternativos, por meio de técnicas computacionais com algoritmos inteligentes, são importantes a fim de aumentar a flexibilidade e a facilidade de adaptação para a contextualização do problema.

A minimização das perdas provocadas por fraudes representa maior possibilidade de investimentos na qualidade do produto, na manutenção do patrimônio da concessão e, principalmente, na possibilidade de melhora no atendimento com menor custo.

Neste contexto, uma ferramenta para a identificação automática de perdas comerciais em sistemas elétricos baseada em algoritmos inteligentes foi proposta.

A partir de uma base de dados disponibilizada por uma concessionária de energia, que foi analisada com diferentes tamanhos de conjuntos de treinamento e teste, com o intuito de calcular a exatidão média para cada variação, foram apresentados os resultados de experimentos utilizando OPF, SVM-RBF, SVM-LINEAR, ANN-MLP e SOM para a classificação de possíveis fraudadores de energia elétrica.

Observou-se que os classificadores OPF e SVM-RBF apresentaram resultados semelhantes no primeiro ensaio, superando os demais classificadores. Entretanto, o OPF foi, aproximadamente, 524 vezes mais rápido que o SVM. No segundo ensaio, o OPF também superou os demais classificadores, tanto na exatidão quanto na soma do tempo de treinamento e classificação para a análise dos perfis de consumidores.

Vale ressaltar, também, que esta pesquisa é pioneira em introduzir o classificador OPF nesta área de pesquisa.

7.2 Trabalhos Futuros

Uma sugestão para os trabalhos futuros é a avaliação de novos classificadores e de novas bases de dados, com diferentes variáveis de entrada para o treinamento dos sistemas inteligentes de maneira a refinar os resultados e melhorar ainda mais o conhecimento e as potencialidades das técnicas inteligentes existentes.

Outra sugestão é o aprimoramento do classificador OPF através de novas abordagens de aprendizagem, já que este se mostrou uma ferramenta inteligente bastante promissora, podendo ser aplicada para solução e minimização de problemas no setor elétrico.

7.3 Publicações Durante a Pesquisa

Durante a execução desta pesquisa foram publicados alguns trabalhos com os resultados preliminares obtidos, e também, trabalhos realizados com o grupo de pesquisa (LSISPOTI), em eventos científicos.

Trabalhos Relacionados com o Tema da Pesquisa

- Lucas I. Pereira, André N. de Souza, Caio C. O. Ramos. **Uma Arquitetura Neural Aplicada na Identificação de Consumidores Fraudadores: Classe Industrial.** *Anais do XX Congresso de Iniciação Científica da UNESP - CIC*, São José dos Campos, 2008.
- Lucas I. Pereira, André N. de Souza, Caio C. O. Ramos. **Técnicas Computacionais Inteligentes na Classificação de Padrões de Consumidores Fraudulentos de Energia.** *Anais do 16º Simpósio Internacional de Iniciação Científica da USP - SIICUSP*, São Paulo, 2008.
- Caio C. O. Ramos, André N. de Souza, Lucas I. Pereira, Danilo S. Gastaldello, João P. Papa. **Técnicas Inteligentes Aplicadas na Identificação de Consumidores Industriais Fraudadores de Energia Elétrica.** *Proceedings of Eighth Latin-American Congress on Electricity Generation and Transmission – CLAGTEE*, Ubatuba, 2009.
- Caio C. O. Ramos, André N. de Souza, João P. Papa, Alexandre X. Falcão. **Fast Power Utility Non-Technical Losses Identification Through Optimum-Path Forest.** *Proceedings of 15th International Conference on Intelligent System Applications to Power Systems – ISAP*, Curitiba, 2009.

Trabalhos Relacionados com o Grupo de Pesquisa (LSISPOTI)

- Caio C. O. Ramos, André N. de Souza, Maria G. Zago. **Estudos de Sobretensões em Linhas Rurais Frente às Descargas Atmosféricas Usando o Atp.** *Anais da VIII Conferência Internacional de Aplicações Industriais – INDUSCON*, Poços de Caldas, 2008.
- André N. de Souza, Caio C. O. Ramos, Maria G. Zago, Pedro da Costa Jr., Jonathan Sunny, Daniel B. Santos. **Aplicação de Ferramentas Computacionais Inteligentes em Estudos de Localização de Falhas em Sistemas Subterrâneos de Distribuição.** *Proceedings of 8th Brazilian Conference on Dynamics, Control and Applications – DINCON*, Bauru, 2009.
- Ricardo Torrezan, Se Un Ahn, André N. de Souza, André C. P. Martins, Caio C. O. Ramos. **Análise Crítica na Aplicação de Ferramentas Computacionais para Identificar e Estimar Problemas Relacionados aos Níveis de Tensão em Sistemas de Distribuição.** *Anais da Conferência Brasileira sobre Qualidade da Energia Elétrica*, Blumenau, 2009.
- André N. de Souza, Paulo S. da Silva, Pedro da Costa Jr., Caio C. O. Ramos. **Algoritmos Computacionais para Quantificação de Erros de Medição Devido a Harmônicos em Sistemas de Medição de Energia.** *Anais da Conferência Brasileira sobre Qualidade da Energia Elétrica*, Blumenau, 2009.
- André N. de Souza, Maria G. Zago, Osvaldo R. Saavedra, Caio C. O. Ramos, Kleber Ferraz. **A Neural Approach to Evaluate the Effect of Lightning in Power Transformers.** *International Journal of Emerging Electric Power Systems*. The Berkeley Electronic Press, vol. 10, iss. 4, article 8, 2009.
- Danilo Sinkiti Gastaldello, André Nunes de Souza, Caio César Oba Ramos, Lucas Imaizumi Pereira. **Desenvolvimento de Ferramentas Computacionais Inteligentes Para Diminuição de Erros em Medidores de Energia Provocados Pelos Harmônicos.** XVII Congresso de Iniciação Científica da UFSCAR – XVII CIC, São Carlos, 2009.
- Danilo Sinkiti Gastaldello, André Nunes de Souza, Caio César Oba Ramos, Alexandre Moriel da Silva. **Implementação de Algoritmos Computacionais para Quantificação de Erros Devido aos Efeitos dos Harmônicos em Sistemas de Medição.** XVII Congresso de Iniciação Científica da UFSCAR – XVII CIC, São Carlos, 2009.
- André N. de Souza, Maria G. Zago, Caio C. O. Ramos, Danilo S. Gastaldello, João P. Papa, Pedro da Costa Jr., José M. Prado, Eurípedes S. Gusmão, Osvaldo R. Saavedra. **Sistema Especialista Integrando Aspectos Elétricos, Estruturais e Ambientais para Diagnóstico da Vida Útil das Famílias de Pára-Raios de 230 KV Instaladas nas Subestações do Sistema da Eletronorte-MT.** *Proceedings of Eighth Latin-American Congress on Electricity Generation and Transmission – CLAGTEE*, Ubatuba, 2009.
- André N. de Souza, Maria G. Zago, Caio C. O. Ramos, Osvaldo R. Saavedra, Shigeaki L. de Lima, Raimundo N. Rosa Filho. **Sistema Especialista Integrando Aspectos Físico-Químicos e Gasosos para o Diagnóstico da Vida Útil de Transformadores de Potência da Eletronorte-MA para Identificação de Falhas em Fase Incipiente.** *Proceedings of Eighth Latin-American Congress on Electricity Generation and Transmission – CLAGTEE*, Ubatuba, 2009.

REFERÊNCIAS

- [1] R. F. Calili, “Desenvolvimento de Sistema para Detecção de Perdas Comerciais em Redes de distribuição de energia elétrica”, Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro-RJ, Julho, 2005.
- [2] E. X. Vieiralses, “Proposta de uma Metodologia para Avaliação das Perdas Comerciais dos Sistemas Elétricos. O Caso Manaus”, Dissertação de Mestrado, Unicamp, Campinas-SP, Agosto, 2005.
- [3] A. C. M. de Araújo, “Perdas e Inadimplência na Atividade de Distribuição de Energia Elétrica no Brasil”, Dissertação de Doutorado, Universidade Federal do Rio de Janeiro, Rio de Janeiro-RJ, Abril, 2007.
- [4] R. Alves, P. Casanova, E. Quirogas, O. Ravelo, W. Gimenez, “Reduction of Non-Technical Losses by Modernization and Updating of Measurement Systems”, *Transmission & Distribution Conference and Exposition: Latin America - TDC '06 IEEE/PES*, pp. 1-5, August, 2006.
- [5] I. E. Davidson, "Evaluation and Effective Management of Nontechnical Losses in Electrical Power Networks", *IEEE AFRICON Conference in Africa*, pp. 473-477, vol.1, October, 2002.
- [6] N. A. Eller, “Arquitetura de Informação para o Gerenciamento de Perdas Comerciais de Energia Elétrica”, Tese de Doutorado, Universidade Federal de Santa Catarina, Florianópolis-SC, Janeiro, 2003.
- [7] CODI – Comitê de Distribuição, “Método para Determinação, Análise e Otimização das Perdas Técnicas em Sistemas de Distribuição”, Relatório CODI 3.2.19.34.0.
- [8] J. A. Bettio, “Constatação de Procedimentos Irregulares e Deficiência no Medidor ou Demais Equipamentos de Medição”, Apostila de Curso, Celesc, Florianópolis-SC, 2001.
- [9] R. M. Queiroga, “Uso de Técnicas de Data Mining para Detecção de Fraudes em Energia Elétrica”, Dissertação de Mestrado, Universidade Federal do Espírito Santo, Vitória-ES, 2005.
- [10] G. V. C. Ortega, “Redes Neurais na Identificação de Perdas Comerciais do Setor Elétrico”, Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro-RJ, Abril, 2008.
- [11] C. J. Bandim, J. E. R. Alves Jr., A.V. Pinto Jr., F.C. Souza, M.R.B. Loureiro, C.A. Magalhães, F. Galvez-Durand, “Identification of Energy Theft and Tampered Meters Using a Central Observer Meter: A Mathematical Approach”, *Transmission and Distribution Conference and Exposition, 2003 IEEE PES*, v.1, pp.163 – 168, Sept., 2003.
- [12] I. H. Witten and E. Frank, "Data Mining: Practical Machine Learning Tools and Techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

- [13] J. Han and M. Kamber, "Data Mining: Concepts and Techniques", Morgan Kaufmann, 2001.
- [14] Z. L Kovacs, "Redes Neurais Artificiais: Fundamentos e Aplicações", Texto Básico Escola Politécnica da Universidade de São Paulo, 2a. Edição, São Paulo, SP, 1996.
- [15] S. Haykin, "Neural Networks – A Comprehensive Foundation", New York: Macmillan, 1994.
- [16] T. Kohonen, "Self-Organizing Map", *Proceedings of the IEEE*, vol.78, pp. 1464-1480, 1990.
- [17] B. Kosko, "Neural Networks and Fuzzy Systems – A Dynamical Systems Approach to Machine Intelligence", Englewood Cliffs: Prentice-Hall, 1992.
- [18] M. H. Hassoun, "Fundamentals of Artificial Neural Networks", MIT Press, Massachusetts, 1995.
- [19] B. Schölkopf and A. J. Smola, "Learning with Kernels", MIT Press, Cambridge, MA, 2002.
- [20] V. N. Vapnik, "An Overview of Statistical Learning Theory", *IEEE Transactions on Neural Networks*, vol. 10, pp. 988-999, 1999.
- [21] B. Schölkopf, R. C. Williamson and P. L. Bartlett, "New Support Vector Algorithms. Neural Computation", MIT Press, pp. 1207-1245, 2000.
- [22] C. Cortes and V. Vapnik, "Support-Vector Networks", *Machine Learning*, AT&T Bell Labs, Kluwer Academic Publishers, pp. 273-297, 1995.
- [23] P. J. Phillips, "Support Vector Machines Applied to Face Recognition", *Advances in Neural Information Processing Systems*, MIT Press, pp. 803-809, 1999.
- [24] G. Guo, S. Z. Li and K. Chan, "An Overview of Statistical Learning Theory", *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 196-201, 2000.
- [25] H. Qiao, S. Zhang, B. Zhang and J. Keane, "Face Recognition Using SVM Decomposition Methods", *International Conference on Intelligent Robots and Systems*, pp. 2015-2020, September, 2004.
- [26] B. Heisele, Y P. Ho and T. Poggio, "Face Recognition with Support Vector Machines: Global Versus Component-Based Approach", *Proc. 8th International Conference on Computer Vision*, pp. 688-694, 2001.
- [27] C. C. Chung and C. J. Lin, "LIBSVM: a Library for Support Vector Machine". <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, March, 2009.
- [28] J. P. Papa, A. X. Falcão, C. T. N. Suzuki, and N. D. A. Mascarenhas, "A Discrete Approach for Supervised Pattern Recognition". *Proc. 12th International Workshop on Combinatorial Image Analysis*, vol. 4958, pp. 136–147, Berlin, 2008.

- [29] J. P. Papa and A. X. Falcão, “A New Variant of the Optimum-Path Forest Classifier”, *Proceedings of the 4th International Symposium on Advances in Visual Computing - ISVC '08*, pp. 935–944, Berlin, Heidelberg, 2008.
- [30] J. P. Papa and A. X. Falcão, “A Learning Algorithm for the Optimum-Path Forest Classifier”, *Proc. 7th IAPR Workshop on Graph-based Representations in Pattern Recognition*, Veneza, Itália, 2009.
- [31] J. P. Papa, A. X. Falcão, P. A. V. Miranda, C. T. N. Suzuki and N. D. A. Mascarenhas, “Design of Robust Pattern Classifiers Based on Optimum-Path Forests”, *Mathematical Morphology and its Applications to Signal and Image Processing (ISMM)*, pp. 337–348. MCT/INPE, 2007.
- [32] A. X. Falcão, J. Stolfi, and R.A. Lotufo, “The Image Foresting Transform: Theory, Algorithms and Applications”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 19–29, 2004.
- [33] R. A. Lotufo and A. X. Falcão, “The Ordered Queue and the Optimality of the Watershed Approaches”, *Mathematical Morphology and its Applications to Image and Signal Processing*, vol. 18, pp. 341–350, Kluwer, Jun., 2000.
- [34] T. M. Mitchell, “Machine Learning”, WCB/McGraw-Hill, 1997.
- [35] R. O. Duda, P. E. Hart, D. G. Stork, “Pattern Classification”, ed. 2, New York: Wiley-Interscience, 2001.
- [36] S. R. Amendolia, “A comparative study of K-Nearest Neighbour, Support Vector Machine and Multi-Layer Perceptron for Thalassemia Screening”, *Chemometrics and Intelligent Laboratory Systems*, vol. 69, issues 1-2, pp. 13-20, Nov., 2003.
- [37] J. Pan, Y. Qiao, S. Sun, “A Fast K Nearest Neighbors Classification Algorithm”, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E87-A, no. 7, Apr., 2004.
- [38] D. E. Goldberg, “Genetic Algorithms in Search, Optimization, and Machine Learning”, Addison Wesley Publishing Company, New York, 1989.
- [39] S. V. Verdú, M. O. Garcia, F. J. G. Franco, N. Encinas, A. G. Marín, , E. G. Lázaro, “Characterization and Identification of Electrical Customers Through the Use of Self-Organizing Maps and Daily Load Parameters”, *Power Systems Conference and Exposition, 2004 IEEE PES*, Oct., 2004.
- [40] Y. M. C Masselli, E. Ferreira, G. L. Torres, L. E. B. da Silva, R. Rossi, L. O. M. dos Reis, “Suggestion of a Neuro-Genetic Hybrid System as an Alternative for the Electric Power Consumption Curves Evaluation”, *Revista Ciências Exatas*, v. 12, pp. 13-18, 2006.
- [41] G. M. F. Paula, “Curvas Típicas de Carga para o Planejamento Operacional do Sistema de Distribuição, Dissertação de Mestrado, Universidade de São Paulo, Curso de Engenharia Elétrica, São Paulo-SP, 2006.

- [42] I. Monederon, F. Biscarri, C. Léon, J. Biscarri, and R. Millán, “Midas: Detection of Non-Technical Losses in Electrical Consumption Using Neural Networks and Statistical Techniques”, *Proceedings of the International Conference on Computational Science and Applications*, vol. 3984, 2006.
- [43] G. G. L. Ferreira, “Desenvolvimento de um Sistema Baseado em Regras para Detecção de Fraudes em Unidades Consumidoras Ligadas em Baixa Tensão, Trabalho de Graduação, Universidade Federal de Santa Maria, Curso de Ciência da Computação, 2007.
- [44] J. E. Cabral, J. O. P. Pinto, E. M. Martins, A. M. A. C. Pinto, “Fraud Detection in High Voltage Electricity Consumers Using Data Mining”, *Transmission and Distribution Conference and Exposition - 2008 IEEE PES*, Apr., 2008.
- [45] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong, S. K. Ahmed, “Non-Technical Loss Analysis for Detection of Electricity Theft Using Support Vector Machines”, *IEEE 2nd International Power and Energy Conference 2008 - PECon 2008*, pp. 907 – 912, Dec., 2008.
- [46] J. Nagi, A. M. Mohammad, K. S. Yap, S. K. Tiong, S.K. Ahmed, “Detection of Abnormalities and Electricity Theft Using Genetic Support Vector Machines”, *IEEE Conference TENCON 2008*, pp. 1 – 6, Nov., 2008.
- [47] J. Koza, “Genetic Programming: on the Programming of Computers by Means of Natural Selection”, The MIT Press, Cambridge, MA, 1992.
- [48] A. H. Nizar, Z. Y. Dong, P Zhang, “Detection Rules for Non-Technical Losses Analysis in Power Utilities”, *IEEE Power Engineering Society General Meeting - Power and Energy Society General Meeting* , pp. 1-8, July, 2008.
- [49] A. H. Nizar, Z. Y. Dong, M. Jalaluddin, M. J. Raffles, “Load Profiling Method in Detecting Non-Technical Loss Activities in a Power Utility”, *IEEE International Power and Energy Conference - PECon '06*, pp. 82–87, November, 2006.
- [50] A. H. Nizar, Z. Y. Dong, J. H. Zhao, P Zhang, “A Data Mining Based NTL Analysis Method”, *IEEE Power Engineering Society General Meeting*, pp. 1-8, June, 2007.
- [51] A. H. Nizar, Z. Y. Dong, Y. Wang, “Power Utility Nontechnical Loss Analysis with Extreme Learning Machine Method”, *IEEE Transactions on Power Systems*, vol. 23, pp. 946-955, Aug., 2008.
- [52] A. H. Nizar, Z. Y. Dong, “Identification and Detection of Electricity Customer Behaviour Irregularities,” *IEEE/PES Power Systems Conference and Exposition 2009 – PES' 09*, pp. 1-10, March, 2009.
- [53] M. E. de Oliveira, D. F. A. Boson, A. Padilha-Feltrin, “A Statistical Analysis of Loss Factor to Determine the Energy Losses,” *IEEE/PES Transmission and Distribution Conference and Exposition: Latin America 2008*, pp. 1 – 6, Aug., 2008.

- [54] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, and C. J. Lin, “LIBLINEAR: A Library for Large Linear Classification,” *Journal of Machine Learning Research*, vol. 9, pp. 1871–1874, 2008.
- [55] João Paulo Papa, Celso Tetsuo Nagase Suzuki, and Alexandre Xavier Falcão, “LibOPF: A Library for the Design of Optimum-Path Forest Classifiers”, 2008. Software version 1.0 available at <http://www.ic.unicamp.br/~afalcao/LibOPF>.
- [56] S. Nissei, “Implementation of a Fast Artificial Neural Network Library (FANN)”, Department of Computer Science University of Copenhagen (DIKU), 2003. Software available at <http://leenissen.dk/fann/>.
- [57] L. I. Kuncheva, “Combining Pattern Classifiers: Methods and Algorithms”, Wiley-Interscience, 2004.
- [58] L. Reyzin and R. Schapire, “How Boosting the Margin Can Also Boost Classifier Complexity”, *Proceedings of the 23th International Conference on Machine learning*, ACM Press, pp. 753–760, New York, NY, 2006.
- [59] R. Collobert and S. Bengio, “Links Between Perceptrons, MLPs and SVMs”, *Proceedings of the 21th International Conference on Machine learning*, ACM Press, pp. 23, New York, NY, 2004.
- [60] N. Panda, E. Chang, and G. Wu, “Concept Boundary Detection for Speeding Up SVMs”, *Proceedings of the 23th International Conference on Machine learning*, ACM Press, pp. 681–688, New York, NY, 2006.